TECHNISCHE UNIVERSITÄT DARMSTADT

TOWARDS A PEACEFUL DEVELOPMENT OF CYBERSPACE

CHALLENGES AND TECHNICAL MEASURES FOR THE DE-ESCALATION OF STATE-LED CYBERCONFLICTS AND ARMS CONTROL OF CYBERWEAPONS

Vom Fachbereich Informatik
der Technischen Universität Darmstadt
genehmigte

DISSERTATION

zur Erlangung des akademischen Grades
Doctor rerum naturalium (Dr. rer. nat.)

vorgelegt von

Thomas Reinhold, Dipl. Inf.
(geboren in Görlitz)

Erstreferent:	Prof. Dr. Dr. Christian Reuter
Korreferent:	Prof. Dr. Volker Roth (Freie Universität Berlin)
Tag der Einreichung:	12. Juli 2023
Tag der Disputation:	15. September 2023
Hochschulkennziffer:	D17

Wissenschaft und
Technik für Frieden
und Sicherheit

Towards a Peaceful Development of Cyberspace - Challenges and Technical Measures for the De-escalation of State-led Cyberconflicts and Arms Control of Cyberweapons
Dissertation, Technische Universität Darmstadt, 2023.

Wissenschaft und Technik für Frieden und Sicherheit (PEASEC)
Fachbereich Informatik
Technische Universität Darmstadt
Jahr der Veröffentlichung: 2023
Tag der mündlichen Prüfung: 2023-09-15
URN: urn:nbn:de:tuda-tuprints-245590
URL: https://tuprints.ulb.tu-darmstadt.de/id/eprint/24559

I

What's so funny about peace, love, and understanding?

— Kettcar

ACKNOWLEDGEMENTS

The dissertation would not have been possible without the huge support of my supervisor *Prof. Dr. Dr. Christian Reuter*. I am thankful for your openness regarding such interdisciplinary topic of arms control, your scientific guidance, your input and support for the numerous papers, your encouragements and for keeping me on track. It has been a great example. I also like to sincerely thank my co-supervisor *Prof. Dr. Volker Roth* (Freie Universität Berlin) for his much appreciated input, advice and ideas on how I can further improve the dissertation.

Thanks a lot also to my colleagues of the research group Science and Technology for Peace and Security (PEASEC) at the Department of Computer Science at the Technical University of Darmstadt, whose help, ideas, feedback and suggestions have helped me a lot to finish this dissertation. A few of the papers were written together with some of you and I thank you for the effort, the excellent cooperation and – in some cases – the shared endurance and patience. I would also like to include in my thanks the (former) student assistants at our research group, especially *Anja-Liisa Gonsior, Hannah Appich, Jenny McMahon* and *Clarissa Neder*, without whose reliable, helpful and meticulous support much of the work would not have been possible.

Furthermore, I would also like to thank all co-authors of my publications (in alphabetical order) *Jonas Franken, Daniel Günther, Dr. Sven Herpig, Philipp Kühn, Helene Pleil, Lilian Reichert, Prof. Dr. Thomas Schneider* and *Dr. Matthias Schulze*. It was a great pleasure to work with you!

I would also like to give special thanks to *Dr. Niklas Schörnig*, whose scientific and political vision and whose humor really helped to keep the track and finish this dissertation.

All this work would not have been possible without the support, advice and wisdom of my wife *Sylvia*. Thank you for the wonderful adventure together. And finally, I am thankful for the continuous support of my parents and friends. I am aware that we always stand on the shoulders of those who came before us. For that I will be forever grateful.

The research work of this thesis has thankfully been funded by the German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the *National Research Center for Applied Cybersecurity ATHENE*, as well as by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) as part of the *Collaborative Research Center CROSSING* (SFB 1119 – 236615297).

ABSTRACT

Cyberspace, already a few decades old, has become a matter of course for most of us, part of our everyday life. At the same time, this space and the global infrastructure behind it are essential for our civilizations, the economy and administration, and thus an essential expression and lifeline of a globalized world. However, these developments also create vulnerabilities and thus, cyberspace is increasingly developing into an intelligence and military operational area – for the defense and security of states but also as a component of offensive military planning, visible in the creation of military cyber-departments and the integration of cyberspace into states' security and defense strategies. In order to contain and regulate the conflict and escalation potential of technology used by military forces, over the last decades, a complex tool set of transparency, de-escalation and arms control measures has been developed and proof-tested. Unfortunately, many of these established measures do not work for cyberspace due to its specific technical characteristics. Even more, the concept of what constitutes a weapon – an essential requirement for regulation – starts to blur for this domain. Against this background, this thesis aims to answer how measures for the de-escalation of state-led conflicts in cyberspace and arms control of cyberweapons can be developed. In order to answer this question, the dissertation takes a specifically technical perspective on these problems and the underlying political challenges of state behavior and international humanitarian law in cyberspace to identify starting points for technical measures of transparency, arms control and verification. Based on this approach of adopting already existing technical measures from other fields of computer science, the thesis will provide proof of concepts approaches for some mentioned challenges like a classification system for cyberweapons that is based on technical measurable features, an approach for the mutual reduction of vulnerability stockpiles and an approach to plausibly assure the non-involvement in a cyberconflict as a measure for de-escalation. All these initial approaches and the questions of how and by which measures arms control and conflict reduction can work for cyberspace are still quite new and subject to not too many debates. Indeed, the approach of deliberately self-restricting the capabilities of technology in order to serve a bigger goal, like the reduction of its destructive usage, is yet not very common for the engineering thinking of computer science. Therefore, this dissertation also aims to provide some impulses regarding the responsibility and creative options of computer science with a view to the peaceful development and use of cyberspace.

CONTENTS

AUTHOR'S PUBLICATIONS

In sum, 41 publications have been published in the context of the author's work.

The following 12 publications are published as chapters of this thesis. Among these publications, the 8 papers in chapters 7, 9, 10, 11, 12, 13, 15 and 16 have been peer-reviewed. The publication in chapter 14 has been written and published in a book (Reinhold & Schörnig, 2022), for which I was involved as co-editor. However, this publication also went through a blind peer-review process. The publications in chapter 6, 8 and 17 are additional papers that haven been originally written for and published in a scientific educational book (Reuter, 2019).

1. Reinhold, T., & Reuter, C. (2019b, March 13). From Cyber War to Cyber Peace. In C. Reuter (Ed.), *Information Technology for Peace and Security: IT Applications and Infrastructures in Conflicts, Crises, War, and Peace* (pp. 139–164). Springer Vieweg. https://doi.org/10.1007/978-3-658-25652-4_7
2. Reinhold, T., & Reuter, C. (2023b). Zur Debatte über die Einhegung eines Cyberwars: Analyse militärischer Cyberaktivitäten im Krieg Russlands gegen die Ukraine. *Zeitschrift für Friedens- und Konfliktforschung*. https://doi.org/10.1007/s42597-023-00094-y
3. Reinhold, T., & Reuter, C. (2019a, March 13). Arms Control and its Applicability to Cyberspace. In C. Reuter (Ed.), *Information Technology for Peace and Security: IT Applications and Infrastructures in Conflicts, Crises, War, and Peace* (pp. 207–231). Springer Vieweg. https://doi.org/10.1007/978-3-658-25652-4_10
4. Reinhold, T., Pleil, H., & Reuter, C. (2023). Challenges for Cyber Arms Control: A Qualitative Expert Interview Study. *Zeitschrift für Außen- und Sicherheitspolitik (ZfAS)*. https://doi.org/10.1007/s12399-023-00960-w
5. Reinhold, T., & Reuter, C. (2021). Towards a Cyber Weapons Assessment Model – Assessment of the Technical Features of Malicious Software. *IEEE Transactions on Technology and Society*, *3*(3), 226–239. https://doi.org/10.1109/TTS.2021.3131817
6. Herpig, S., & Reinhold, T. (2018, October). Spotting the Bear: Credible Attribution and Russian Operations in Cyberspace. In N. Popescu & S. Secrieru (Eds.), *Hacks, leaks and disruptions: Russian cyber strategies* (pp. 33–42, Vol. 148). European Union Institute for Security Studies (EUISS). https://www.jstor.org/stable/resrep21140.7
7. Schulze, M., & Reinhold, T. (2018). Wannacry About the Tragedy of the Commons? Game-Theory and the Failure of Global Vulnerability Disclosure. *European Conference on Information Warfare and Security, ECCWS*, *2018-June*, 454–463. https://www.proquest.com/openview/f6ccddd62973bd8997c3fcd40951f4f1/1?cbl=396497&pq-origsite=gscholar&parentSessionId=7jm9tc94UMKaTk9pAtTjzd%2BhJYdl8V55qGHqrUpnUM8%3D

8. Franken, J., Reinhold, T., Reichert, L., & Reuter, C. (2022). The Digital Divide in State Vulnerability to Submarine Communications Cable Failure. *International Journal of Critical Infrastructure Protection*, *38*. https://doi.org/10.1016/j.ijcip.2022.100522

9. Reinhold, T., & Reuter, C. (2022, October 9). Cyber Weapons and Artificial Intelligence: Impact, Influence and the Challenges for Arms Control. In T. Reinhold & N. Schörnig (Eds.), *Armament, Arms Control and Artificial Intelligence: The Janus-faced Nature of Machine Learning in the Military Realm* (pp. 145–158). Springer International Publishing. https://doi.org/10.1007/978-3-031-11043-6_11

10. Reinhold, T., & Reuter, C. (2023a). Preventing the Escalation of Cyber Conflicts: Towards an Approach To Plausibly Assure the Non-Involvement in a Cyberattack. *Zeitschrift für Friedens- und Konfliktforschung (ZeFKo)*. https://doi.org/10.1007/s42597-023-00099-7

11. Reinhold, T., Kühn, P., Günther, D., Schneider, T., & Reuter, C. (2023). EXTRUST: Reducing Exploit Stockpiles With a Privacy-Preserving Depletion System for Inter-State Relationship. *IEEE Transactions on Technology and Society*. https://doi.org/10.1109/TTS.2023.3280356

12. Reinhold, T., & Reuter, C. (2019c, March 13). Verification in Cyberspace. In C. Reuter (Ed.), *Information Technology for Peace and Security: IT Applications and Infrastructures in Conflicts, Crises, War, and Peace* (pp. 257–275). Springer Vieweg. https://doi.org/10.1007/978-3-658-25652-4_12

The following 29 publications are not included in the thesis. These include scientific papers, journalistic contributions from the field of peace and conflict research in the natural sciences, as well as publications with a strong political focus. Nevertheless, their findings are supplementary to this thesis and represent the self-entitlement of this thesis to address socially and politically relevant topics and questions as a computer scientist.

13. Reinhold, T. (2013). Malware als Waffe. *ADLAS Magazin für Außen- und Sicherheitspolitik*, *13*(1), 7–11. https://adlasmagazin.files.wordpress.com/2013/02/adlas-0113.pdf

14. Reinhold, T. (2014c). Internationale Kooperationsrichtlinien – ein Ausweg aus dem Attributionsdilemma. *Sicherheit und Frieden (S+F) / Security and Peace*, *32*(1), 23–27. https://www.jstor.org/stable/24233888

15. Reinhold, T. (2014b). Friedens- und Sicherheitspolitische Fragen zur Militarisierung des Cyberspace. *FIfF Kommunikation*, (4), 70–73. http://www.fiff.de/publikationen/fiff-kommunikation/fk-2014/fk-2014-4

16. Reinhold, T. (2015a). Militarisierung des Cyberspace – Friedens- und Sicherheitspolitische Fragen. *Wissenschaft und Frieden*, (2), 31–34. https://wissenschaft-und-frieden.de/artikel/militarisierung-des-cyberspace/

17. Reinhold, T. (2015c). Betrifft: Cyberpeace – Auswirkungen der Exportbeschränkungen von Cyberwaffen durch das Wassenaar-Abkommen. *FIfF Kommunikation*, (1), 6–7. https://www.fiff.de/publikationen/fiff-kommunikation/fk-2015/fk-2015-1/fk-2015-1-content/fk-1-15-s6.pdf

18. Reinhold, T. (2015b). Möglichkeiten und Grenzen zur Bestimmung von Cyberwaffen. In Cunningham, Douglas, Hofstedt, Petra, Meer, Klaus, & Schmitt, Ingo (Eds.), *INFORMATIK 2015* (pp. 587–596, Vol. 246). Gesellschaft für Informatik e.V. http://dl.gi.de/handle/20.500.12116/2219

19. Reinhold, T. (2016b). Der Cyberspace – Vorfälle, militärische Aufrüstung und erste Friedensbestrebungen. *Welttrends – das außenpolitische Journal*, *113*(3), 22–27

20. Reinhold, T. (2016c). Die Bundeswehr zieht ins Cyberfeld – Ein Kommentar zum Aufbau des neuen Bundeswehr-Organisationsbereiches Cyber- und Informationsraum. *Blätter für deutsche und internationale Politik*, 17–20. https://www.blaetter.de/archiv/jahrgaenge/2016/juli/die-bundeswehr-zieht-ins-cyberfeld

21. Reinhold, T. (2016d). Cyberspace als Kriegsschauplatz? Herausforderungen für Völkerrecht und Sicherheitspolitik“ (APUZ 35–36/2016). *Aus Politik und Zeitgeschichte (APUZ), Bundeszentrale für politische Bildung*, *35–36*. http://www.bpb.de/apuz/232966/cyberspace-als-kriegsschauplatz

22. Reinhold, T., & Schulze, M. (2017). Digitale Gegenangriffe Eine Analyse der technischen und politischen Implikationen von „hack backs“. *SWP Arbeitspapier*, *01.08.2017*(1), 1–18. https://www.swp-berlin.org/publications/products/arbeitspapiere/AP_Schulze_Hackback_08_2017.pdf

23. Reinhold, T. (2016a). Zur Verantwortung der Informatik in einer technologisierten Gesellschaft. *Sicherheit und Frieden*, *34*(4), 253–256. https://doi.org/10.5771/0175-274X-2016-4-253

24. Reinhold, T. (2018c). Rethinking the Attribution Problem – A Plausible Proof of Non-Involvement as an Alternative to Credible Attribution [magazine]. *Issue brief 2: Briefing and memos from the research advisory group of the Global Commission on the Stability of Cyberspace (GCSC)*, (2), 134–149. https://hcss.nl/wp-content/uploads/2022/06/GCSC-Research-Advisory-Group-Issue-Brief-2-Bratislava.pdf

25. Reinhold, T. (2019b). Rüstungskontrolle für den Cyberspace – Herausforderungen und erste Ansätze [magazine]. *FIfF-Kommunikation 3/2019 "Cyberpeace und IT-Security"*, (3), 26–29. https://www.fiff.de/publikationen/fiff-kommunikation/fk-jhrg-2019/FK-2019-3

26. Reinhold, T. (2019c, December 12). Cyberspace as Military Domain: Monitoring Cyberweapons. In D. Feldner (Ed.), *Redesigning Organizations: Concepts for the Connected Society* (pp. 267–278). Springer International Publishing. https://doi.org/10.1007/978-3-030-27957-8_20

27. Reuter, C., Riebe, T., Aldehoff, L., Kaufhold, M.-A., & Reinhold, T. (2019). Cyberwar zwischen Fiktion und Realität – technologische Möglichkeiten. In I.-J. Werkner & N. Schörnig (Eds.), *Cyberwar – die Digitalisierung der Kriegsführung: Fragen zur Gewalt. Band 6* (pp. 15–38). Springer VS. https://doi.org/10.1007/978-3-658-27713-0_2

28. Reinhold, T. (2019a). Counting Cyber Weapons – New Approaches to Regulate and Control Destructive Cyber Tools. *SCIENCE PEACE SECURITY ’19 - Proceedings of the Interdisciplinary Conference on Technical Peace and Security*

Challenges, 41–45. https://tuprints.ulb.tu-darmstadt.de/9164/2/2019_SciencePeaceSecurity_Proceedings-TUprints.pdf

29. Reinhold, T. (2018b). Der Cyberspace ein neuer Kriegsschauplatz? Herausforderungen für Völkerrecht und Sicherheitspolitik. *Der Auftrag, Verbandszeitschrift der Gemeinschaft Katholischer Soldaten (GKS)*, *302*(2), 30–35. https://www.gemeinschaft-katholischer-soldaten.de/attachments/article/104/Auftrag_302_PDF_W.pdf

30. Riebe, T., Kaufhold, M.-A., Kumar, T., & Reuter, C. (2019). Threat Intelligence Application for Cyber Attribution. In C. Reuter (Ed.), *SCIENCE PEACE SECURITY '19 - Proceedings of the Interdisciplinary Conference on Technical Peace and Security Challenges* (pp. 56–59). TUprints. https://tubiblio.ulb.tu-darmstadt.de/116000/

31. Reinhold, T. (2020b, December 3). *Stellungnahme zur öffentlichen Anhörung im Verteidigungsausschuss des deutschen Bundestages zum Thema: Verfassungs- und völkerrechtliche Fragen im militärischen Cyber- und Informationsraum unter besonderer Berücksichtigung des Parlamentsvorbehalts, der Zurechenbarkeit von Cyberangriffen sowie einer möglichen Anpassung nationaler und internationaler Normen*. Deutscher Bundestag, Verteidigungsausschuss. https://www.bundestag.de/resource/blob/824622/67fc9db4f856a8445355562500d2a134/stellungnahme-Thomas-Reinhold_15-03-2021-data.pdf

32. Reinhold, T. (2020a, June 9). *Russian Hacker Wanted!* Directions Blog. https://directionsblog.eu/russian-hacker-wanted/

33. Reinhold, T. (2021d). Export Controls on Cyber-Surveillance Items. *Newsletter of the EU Non-Proiliferation and Disarmament Consortium*, (33), 1. https://www.iai.it/sites/default/files/eunpd_e-newsletter_33.pdf

34. Reinhold, T. (2021c). Zur Rolle und Verantwortung der Informatik für die Friedensforschung und Rüstungskontrolle. *FIfF Kommunikation*, (4), 47–49. https://www.fiff.de/publikationen/fiff-kommunikation/fk-2021/fk-2021-4/fk-2021-4-content/fk-4-21-p47.pdf

35. Reinhold, T. (2021a). Der Weg zu einem sicheren zivilen Cyberspace. *Hoch3 - Die Zeitung der Technischen Universität Darmstadt*, *17*(10), 14. https://www.tu-darmstadt.de/universitaet/aktuelles_meldungen/publikationen/publikationen_archiv/einzelansicht_12672.de.jsp

36. Reinhold, T. (2021b). Überlegungen zur Militarisierung Künstlicher Intelligenz - Von Fallstricken, Grenzen und Problemen der Rüstungskontrolle. *Wissenschaft und Frieden - Dossier*, *4*(93). https://wissenschaft-und-frieden.de/dossier/kuenstliche-intelligenz-zieht-in-den-krieg/

37. Reinhold, T. (2021e, December 30). *Export Control of Surveillance Software from Germany and Europe - Regulations, Limits and Weaknesses*. Heinrich-Böll-Stiftung. https://il.boell.org/en/2021/12/27/export-control-surveillance-software-germany-and-europe-regulations-limits-and

38. Reinhold, T., & Schörnig, N. (Eds.). (2022). *Armament, Arms Control and Artificial Intelligence: The Janus-faced Nature of Machine Learning in the Military Realm* (1st ed.). Springer Cham. https://doi.org/10.1007/978-3-031-11043-6

39. Schörnig, N., & Reinhold, T. (2022). Introduction. In T. Reinhold & N. Schörnig (Eds.), *Armament, Arms Control and Artificial Intelligence: The Janus-faced Nature of Machine Learning in the Military Realm* (pp. 1–9). Springer International Publishing. https://doi.org/10.1007/978-3-031-11043-6_1

40. Reinhold, T. (2022). Arms Control for Artificial Intelligence. In T. Reinhold & N. Schörnig (Eds.), *Armament, Arms Control and Artificial Intelligence: The Janus-faced Nature of Machine Learning in the Military Realm* (pp. 211–226). Springer International Publishing. https://doi.org/10.1007/978-3-031-11043-6_15

41. Reuter, C., Riebe, T., Haunschild, J., Reinhold, T., & Schmid, S. (2022). Zur Schnittmenge von Informatik mit Friedens- und Sicherheitsforschung: Erfahrungen aus der interdisziplinären Lehre in der Friedensinformatik. *Zeitschrift für Friedens- und Konfliktforschung*, *11*(2), 129–140. https://doi.org/10.1007/s42597-022-00078-4

SYNOPSIS

1 INTRODUCTION

1.1 Motivation and Problem Statement

Cyberspace, understood as the entirety of IT network systems, their hardware and software infrastructure, as well as the information contained, processed and transmitted therein (NIST, 2012), is increasingly becoming an intelligence and military operational area (Faesen et al., 2022; Kai et al., 2022). States are creating military cyber departments and integrating this domain into their defense policy strategies (Kastelic et al., 2021) and Russia's attack on Ukraine in 2022 presented a first glimpse of the role which cyberspace activities of state and non-state actors can play in an actual armed conflict (FortiGuard, 2023; Greenberg, 2023). The militarization of cyberspace, a virtual, non-physical domain that defies the rules of physical space, poses new challenges for established approaches of sustaining international stability (Futter, 2020). Even if important political steps have been initiated to develop a common international perspective on these challenges (German Government, 2021; Security & Europe, 2016), actual measures that can regulate or even limit this development in cyberspace are still missing. Comparable technological transformation processes also referred to as a revolution in military affairs (RMA) have occurred repeatedly in the past decades, such as the development of nuclear weapons and missile technology, bio- and chemical weapons, as well as the efforts to weaponize outer space. In order to limit the military escalation potential of such developments, various measures have been established and further developed over the years. On the one hand, these were intended to safeguard national security interests and, on the other hand, to create inter-state stability. One of these measures has been arms control and disarmament, that played quite an important role with its dualism of political as well as practical approaches:

> Arms control (...) refers to mutually agreed upon restraints or controls (usually between states) on the development, production, stockpiling, proliferation, deployment and use of troops, small arms, conventional weapons and weapons of mass destruction. Arms control includes agreements that increase the transparency of military capabilities and activities, with the intention of reducing the risk of misinterpretation or miscalculation. (NATO, 2023)

Unlike, for example, companies, which are bound by national or international legislation, state behavior is usually based on voluntariness and the willingness to cooperate and comply with committed rules. This means that arms control measures cannot prevent states from performing activities that are contrary to agreed behavior or from gray area actions like espionage. Arms control negotiation usually accepts these limitations in order to achieve at least any kind of agreement that can help to mitigate the military

threats, which historically worked best when states had mutually shared interests. An essential, yet not always included, component of arms control agreements that addresses this aspect is the accompanying implementation of specific procedures that enable states to mutually control the contractual commitments, the so-called verification regimes. Such procedures are usually practical, down-to-earth measures of counting things, measuring specific values and comparing them with defined thresholds or surveilling and monitoring industrial facilities[1] (Woolf, 2010). Especially for these parts of arms control treaties, peace and conflict research within the natural sciences has made significant contributions over the last decades by providing necessary technical knowledge, defining rules and limits that reflect the desired degree of control and by developing technical means for monitoring, measuring and compliance checking. As computer scientists are the experts that understand and shape the technical foundations of cyberspace as well as create the software that runs everything within this domain, the challenges and threats that arise from the militarization of cyberspace highlights their responsibility regarding the peaceful development of this domain (Reinhold, 2016a; Reuter, Aldehoff, et al., 2019) or, as Jürgen Altmann put it:

> There is a particular gap in research on IT-specific issues. Synergies with the civilian IT security and network monitoring measures already taking place can certainly play a major role. (Altmann, 2019b) (*translated from German*)

The following thesis aims to take these demands seriously by using the tools, research methods and technological concepts of computer science and especially cybersecurity to analyze the political requirements and needs of arms control and de-escalation, to gain insights into the technical aspects of these challenges and to provide measures and approaches towards their solution. Connected with this approach is the hope and desire, to provide impulses for connections between the computer science and peace and conflict research and to inspire further research on how technical solutions can help sustain a peaceful cyberspace.

1.2 Related Work and Research Gap

Regarding early arms control and de-escalation measures for cyberspace, some work has already been done. One popular, yet important approach is the proposal to map the lessons learned from other militarized technologies to cyberspace, in order to assess the specific challenges of cyberspace but also the chances to adopt already existing arms control measures. Hansel et al. (2018) have done this for preventive arms control of ballistic missiles, Kühn (2014) for tactical ballistic missiles (TCBM), Maybaum and Tölle (2016) analyzes arms control of conventional military technologies Maybaum and

[1] Examples of such verification regimes are the Treaty on Open Skies (OSCE, 1992), which allows a state to fly over the territory of the contracting parties on defined routes and to take photographs, radar and infrared images in order to check military installations and stocks. Other examples are the international sensor network of the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) for seismic tremors and the network of atmospheric radionuclide detectors, which can be used to detect underground nuclear tests. Another, quite prominent example is the organization of the UN-affiliated International Atomic Energy Agency (IAEA), whose tasks include monitoring Iran's nuclear program under the JCPOA ("Joint Comprehensive Plan of Action") treaty, for example by means of on-site inspections (IAEA, 2015b).

Tölle (2016) and Litwak and King (2015) assesses the experiences from nuclear arms control. All of these approaches come to the conclusion that cyberspace possesses some unique features, that inherently differ from physical domains, hindering the transfer of established approaches to this domain. Some experts even argue that these aspects, as well as the technical principles of cyberspace, render these attempts practically unimplementable or completely impossible, such as Burgers and Robinson (2018) and Perkovich and Levite (2017). Other work, for example by Altmann (2019a) and Pawlak (2016) or proposals like Security and Europe (2016) confirm these fundamental problems, arguing for the necessity to develop a common international understanding of the topic and the threats and dependencies of cyberspace. Further work focuses on the question of how malicious cyber tools could actually be used in an open conflict or declared war, and how this might differ from the omnipresent espionage activities. Current perspectives range from a primarily intelligence-focused meaning of cyberweapons as proposed by Rovner (2020) on one end, to the usage of cyberweapons as a replacement for conventional warfare and boots on the ground as argued by Bigelow (2019) on the other. In connection with these considerations, further research like by Buchanan and Cunningham (2020) discusses how the complex, usually long-ranging processes of military planning, that involves comprehensive preparation, signal intelligence, reconnaissance etc. will shape the activities of military forces in this domain.

Nevertheless, despite all these challenges, cyberspace is a completely human-designed and continuously adapted domain (Shea, 2018), and the further development of its functional principles is entirely possible. This opens possibilities for the development of conflict de-escalation and arms control measures for cyberspace Reuter et al. (2022), but also requires the adoption or even the reinvention of concepts and definitions that build the foundation for arms control. This includes specific challenges like the central question of what constitutes a cyberweapon as raised by Biller and Schmitt (2019), or how the strong dual use character of cyberspace can be taken into account, as discussed by Riebe and Reuter (2019a, 2019b). Other work, like Stevens (2018), focuses on the challenge of how regulation, monitoring and dismantling can work for something that does not even physically exist and the fundamental impossibility of quantifying software digital data as argued by Hansel and Nanni (2018), which counteracts control procedures based on the limitation and quantity determination as outlined by Eilstrup-Sangiovanni (2018) that are essential for verification measures. Additional aspects that directly relate to this characteristic of cyberspace are addressed by J. A. Silomon (2020) who highlights the uncontrollable spread of malware and the knowledge of security vulnerabilities, by UNHCR (2019) that analyses the problems with the traceability of the transfer of defense equipment and defense technologies or by Bendiek and Schulze (2021) and Broeders et al. (2020) who examine the difficulties in attributing cyberattacks.

The mentioned work provided highly required and valuable input for the national as well as international debates regarding the militarization of cyberspace and how to mitigate the effects of this development. Nevertheless, as far as it is possible to judge these different kind of works together, they have in common that they usually address political levels regarding security policies and are therefore based on political science and its approaches, theories, and concepts. They therefore do not analyze the specifics of cyberspace on the technical level of the hard- and software that shape the cyberspace and its infrastructures, thus failing to provide practically applicable measures for arms control or de-escalation of conflicts in cyberspace. This shortcoming is especially exemplified by the lack of concepts for assessing potentially harmful software as

a cyberweapon. Current research either proposes an assessment of the triggered effects and its comparison to effects of conventional weapons (e.g. CCDCOE (2017) and Herr (2014)) or an assessment of the presumable intent or the strategic considerations of the attacker (e.g. Dewar (2017) and Orye and Maennel (2019)). This approach is not applicable for a regulation in the context of arms control, where an assessment has to take place *before* the actual usage of an assessed tool, leaving out any speculations about a possible intent or strategic goal of a hypothetical attacker. This essential requirement for any regulation has to be based upon directly measurable capabilities and by assessing the technical characteristics of an analyzed tool, being the only available parameters. So far, no research has taken this into account. Existing approaches, for example from Hatch (2018) and Maathuis et al. (2016), are usually hindered by the lack of a fundamental analysis of the technical peculiarities of weaponized malicious cyber tools in contrast to other weapons and the physical domains that they are aimed at, like sea, air and land. A similar situation exists for the proposal of practical disarmament or de-escalation measures in cyberspace. Whereas organizational proposals for a vulnerability equity process (VEP) that should mitigate the negative effects of states stockpiling vulnerabilities, as made for example by Schulze (2019) and US White House (2017), no concepts have been developed so far that would allow the reduction of vulnerability stockpiles of multiple states as a disarmament approach. Such measure would require any kind of comparison capability between stockpiles, and an algorithm that takes the technical properties of vulnerabilities and their machine-readable description into account. So far, this has not been proposed. Similarly, this lack of technical-rooted concepts also affects the discussed challenge of verification approaches, where measuring technical properties lies at its core. The same holds true for the de-escalation of imminent or ongoing conflicts in cyberspace, with its necessity to mitigate threats arising from false-flag operations by either uncovering the actual attacker or otherwise provide information about the non-involvement into an attack.

1.3 Aim and Research Question

In light of the identified research gap, this thesis aims to contribute to the development of practical cyber arms control and de-escalation measures, that can actually be implemented and applied alongside political efforts for the peaceful development of this domain. With this regard, the main research question of this thesis is:

How can measures for the de-escalation of state-led conflicts in cyberspace and arms control of cyberweapons be developed?

In order to answer this question, the following sub-questions are required to understand the specific constraints, identify starting points for arms control in cyberspace and to finally develop such measures. The methodological approach of consecutively answering these question is further explained in chapter 1.4.

- Which domain-specific characteristics of cyberspace make it difficult to transfer established procedures for verification measures of other, well-researched weapons-capable technologies such as nuclear, biological, chemical and conventional weapons to cyberspace and why?

This question and its different aspects are dealt with in particular in the chapters 6, 7, 8 and 9.

- What are the challenges that arise from malicious state-led activities in cyberspace, its militarization and current technological trends that undermine the security of this domain, which can be addressed by arms control and de-escalation measures? This question and its different aspects are dealt with in particular in the chapters 11, 12, 13 and 14

- Which existing procedures from other areas of computer science can be adopted and applied to develop arms-control and de-escalation measures in cyberspace? What are suitable technical parameters, technical requirements and limitations for implementing such measures?
  This question and is dealt with in particular in the chapters 15, 16 and 17.

This work is intended as a basis for further concrete implementations of arms control and de-escalation procedures in cyberspace. The developed and presented measures should provide a thought-provoking perspective, that cyber arms control can actually work and how such measures can technically be implemented. Nevertheless, this work can only be a start to this complex topic. It therefore aims to provide impulses for further scientific peace research with its transfer of concepts and methods of arms control, dual-use assessment, non-proliferation and disarmament to cyberspace.

Finally, it is important to highlight that the presented measures of arms control are only intended for IT systems of institutions which are already under direct state legislation or control and likely to be responsible for cyber activities in foreign IT networks, such as military networks or the IT systems and networks of intelligence services. While this has its limitations, it prevents the establishment of unlawful surveillance measures, which would ultimately contradict the purpose of fostering a peaceful development of cyberspace.

1.4 Methodology and Structure of the Thesis

This dissertation consists of four parts: a synopsis and three parts with the publications.

Synopsis

Chapter 1 presents the motivation for developing IT-based measures for the de-escalation of state-led conflicts and arms control for cyberspace (1.1). It describes the context and the related work of this dissertation (1.2), outlines the aims and objectives (1.3), summarizes the structure of the document (1.4) and presents the containing publications together with the authors contributions (1.5).

Chapter 2, 3 and 4 present the steps that have been undertaken to answer the different research questions and the empirical and theoretical findings of this thesis.

Chapter 5 presents the discussion, the limitations and conclusion of this dissertation with regard to the research question as well as perspectives for future work.

Publications

The publications that are published as chapters of this thesis are structured into three thematic parts that reflect the methodical approach of this work. In the following, the content, methodology and contribution of each of these parts are presented, with a focus on their role in answering the central research question of this dissertation. A detailed presentation of all publications is given in chapter 1.5.

Part A – Concepts and Challenges of Peace in Cyberspace: Regarding the identified research gap and the interdisciplinary nature of this work between computer science and political requirements, a first necessary step is to analyze what arms control means for cyberspace and what the challenges are for implementing arms control in this domain. This helps to identify the constraints, limitations and specific boundary conditions for the development of cyber arms control and cyberconflict deescalation measures and creates a foundation for the following thematic parts B and especially part C, which contain the actual results of this thesis.

The first thematic part therefore starts with discussing the different aspects of peace, war, and the militarization of cyberspace as a point of reference and for setting relevant definitions. Building up on this, the publications of this part focus on the questions above by firstly assessing the existing approaches for other weaponized technologies and how they can be applied to the domain cyberspace (6). Especially in light of the Russian war against Ukraine, it then discusses what the term *war* means for conflicts in cyberspace and how the events seen in Ukraine match the so far existing scholarly perspectives of a *cyberwar* (7). Following up on this, it then analyzes the technical specificities of the cyberspace domain that prevent establishing arms control approaches (8) and assesses which further challenges towards the establishment of such measures exist (9).

The methodological approaches of the publications collected in part A are based primarily on cross-sectional literature analyses between information technology and, in particular, IT security as well as political science with a focus on security policy (6, 7, 8, 9). In addition, the literature reviews are supported by qualitative expert interviews (9) to assess the challenges for a cyber arms treaty.

Among peer-reviewed publications (7 and 9), part A also contains two papers (6 and 8) that were originally written for and published in a scientific educational book (Reuter, 2019). Due to the "leveling the ground" character of this first part, these papers are included, as they provide an overview of the mentioned challenges and the necessary terminology.

Part B – Threats From Malicious Activities in Cyberspace and Technological Trends: Based on part A, the next step towards the reduction or prevention of state based cyberconflicts is the identification of the specific threats for and in cyberspace by malicious activities of states and how they differ from threats in physical domains.

Regarding the potential regulation of the weaponized malicious cyber tools, this especially regards the analysis of what a state-owned *cyberweapon* actually is and how software can be evaluated and classified as such. The assessment further has to include challenges that arise from technological trends, that already or may potentially influence the cyberspace domain and how malicious activities are or might be performed. These insights allow for the identification of starting points for measures of arms control in cyberspace and provide a dedicated overview of the specific problems that have to be addressed for a peaceful development of cyberspace.

To realize this objective, the second thematic part, as a first step, develops a model for the definition of cyberweapons, that didn't require speculations about the presumable attacker or their intent and that therefore can be applied before the actual usage of a malicious cyber tool (10). Further on, the attribution problem, as an all-affecting challenge of cyberspace is analyzed, to understand its influence on the escalation dynamics of conflicts in cyberspace and the risks of conflicts by mistake (11). After that, the problem of vulnerability and exploit stockpiling, as the *"base material"* is addressed, by discussing how and under what conditions a formal mechanism on a national level (a so-called vulnerability equity process - VEP) can support the reduction of such stockpiles as a measure of disarmament (12). In addition to conflicts *in* cyberspace, the very infrastructures that constitute this domain itself are increasingly becoming part of conflicts. Therefore, the vulnerabilities and dependencies of the global undersea fiber optic infrastructures are analyzed (13), to assess where arms control and de-escalation measures need to include the physical space. Finally, in order to include potential upcoming challenges, the current overarching technological disruptive trend of artificial intelligence is assessed to take its influence on the weaponization of cyberspace into account (14).

Methodologically, the publications of part B also use a cross-sectional literature analysis to analyze the current research perspectives on the mentioned challenges (10, 11, 12, 13, 14). Beyond that, further approaches have been applied. The cyberweapon assessment (10) defines a framework that is built upon a step-by-step analysis of multiple technical measurable parameters that follow the life cycle of weapons, alongside the respective assessment spectrum for each parameter. The vulnerability disclosure assessment (12) uses a game-theory approach, that applies the theoretic concept of a *prisoners' dilemma* to colliding national security interests, which results in what game theory calls a tragedy of the commons, to develop best practices for escaping a situation where rational-actions on a local level produce irrational effects on a global scale. Finally, the global undersea fiber optics network is modelled based on graph theory concepts (13) and analyzed with quantitative network analysis to quantify redundancies as a measure of dependency and vulnerability for territorial entities.

The publication in chapter 14 has been written and published in a book (Reinhold & Schörnig, 2022), for which I was involved as co-editor. However, this publication also went through a blind peer-review process.

Part C – Approaches for the peaceful development of cyberspace: The final step concludes the previous parts with the presentation and discussion of actual approaches for the peaceful development of cyberspace. It develops and evaluates technical procedures for de-escalation in cyberconflicts as well as for the disarmament of cyberweapons in order to show that and how already existing computer science research can be used and

adapted for this task. Given the complex nature of this topic, the possible application of the presented approaches and the necessary technical adjustments are discussed, as well as the limitations and the required political steps of involved actors.

With regard to this goal and the underlying research question, the third part first develops a technical concept of how to collect network connection information, that can plausibly assure the non-involvement of an actor in a cyberconflict (15) in order to mitigate the escalation potential of false-flag operations. With regard to the non-disclosure of vulnerabilities and exploits, a second technical concept called ExTRUST presents an approach, describing how state actors can compare their stockpiles without revealing any secrets or the necessity to uncover the actual vulnerabilities (16). This allows participating actors to detect identical items that can be considered for disclosure as a measure of disarmament. The third paper analyzes the possibilities and limitations of verification measures for malicious cyber capabilities (17). It presents technical approaches and starting points for state actors to mutually control and oversee each others "military cyber arsenals" as an important pillar of potential arms control treaties.

Regarding the methodology, the publications of part C develop technical concepts (15 and 17) or fully implementable algorithms (16). A core approach of these works has been the focus on already existing IT measures and approaches, especially from IT security and cryptography. The presented concepts therefore analyze and discuss the usage of existing measures as well as its limitation for arms control and necessary adoption or extension.

Part C also contains a paper (17 that was originally written for and published in a scientific educational book (Reuter, 2019). This paper has been included as it analyses the broad range of practical arms control approaches and assesses further possible starting points for applying them to cyberspace. Together with the presentation of the developed technical measures (15 and 16), this hopefully presents a good basis for future work.

1.5 Underlying Publications and Contributions of the Authors

In the following, all publications that are combined in the three thematic blocks and which have been previously published as articles (journal articles, conference papers and book chapters) are briefly presented alongside their publication reference. The publications have been created by multiple authors, with me in the role of the leading and corresponding author. However, the contributions of the co-authors, which have been gratefully provided, are very important as well. Therefore, the independent academic contributions to each publication, as confirmed by all authors, will be declared for each chapter. External co-authors that are not part of the research group "Science and Technology for Peace and Security" (PEASEC) at the Department of Computer Science at the Technical University of Darmstadt are listed with their respective affiliation.

Please note: For better readability, the spelling of the terms "*cyberspace*", "*cyberattack*", "*cybersecurity*", "*cyberpeace*", "*cyberwar*", "*cyberwarfare*", "*cyberweapon*" as well as "*cyberconflict*" was standardized and adapted accordingly in the publications. The spelling in the respective publication references has not been changed. Additionally,

the spelling of the publication texts has been unified to American English and spelling mistakes have been corrected.

1.5.1 *Part A: Concepts and Challenges of Peace in Cyberspace*

Chapter 6: "From Cyberwar to Cyberpeace" analyzes the ongoing trend and the security implications of the militarization of cyberspace based on historical developments and transformations due to advancements in military technologies as well as the political progress that has been made and peace preserving political tools that have been developed since. It discusses the challenges of applying these established measures to cyberspace and emphasizes the role of social initiatives, such as the cyberpeace campaign of the "Forum InformatikerInnen für Frieden und gesellschaftliche Verantwortung" (Forum of computer scientists for peace and societal responsibility).

Published as: Reinhold, T., & Reuter, C. (2019b, March 13). From Cyber War to Cyber Peace. In C. Reuter (Ed.), *Information Technology for Peace and Security: IT Applications and Infrastructures in Conflicts, Crises, War, and Peace* (pp. 139–164). Springer Vieweg. https://doi.org/10.1007/978-3-658-25652-4_7

Author Statement Thomas Reinhold has written the chapters 6.1 to 6.4, performed the necessary literature analysis as well as the assessment of the state-of-the-art perspective on cyberwar challenges. Christian Reuter supported the refinement of the argumentation and structuring of the paper.

Chapter 7: "Military Cyber Activities in Russia's War Against Ukraine and Their Significance for the Debates on the Containment of a Cyberwar" analyzes the military activities in cyberspace in the context of Russia's war against Ukraine and assesses these events regarding the previously prevailing notion and perspectives of what a cyberwar could be. Based on this, conclusions are drawn, firstly regarding the future significance of cyber activities for Russia in times of peace and conflict in terms of the general military use of cyber assets, and secondly with regard to future international debates on the containment of cyberwar and the harmful use of cyber assets. The included text is a translation of the original German publication.

Published as: Reinhold, T., & Reuter, C. (2023b). Zur Debatte über die Einhegung eines Cyberwars: Analyse militärischer Cyberaktivitäten im Krieg Russlands gegen die Ukraine. *Zeitschrift für Friedens- und Konfliktforschung*. https://doi.org/10.1007/s42597-023-00094-y

Author Statement Thomas Reinhold has written the chapters 7.1 to 7.4 and performed the underlying analysis of the events during the war, based on media news, as well as the following initial scientific analyses. Thomas Reinhold has performed the comparison with prior scientific approaches towards the concept cyberwar and performed the conclusion towards expectable further malicious events in cyberspace in progress of this war, as well as possible changes in the perception of the concept of "cyberwar" going beyond it. Christian Reuter

supported the refinement of the argumentation, structuring of the paper and with providing insights into the scientific perspectives of cyberwar and cyberpeace.

Chapter 8: "Arms Control and its Applicability to Cyberspace" discusses the necessity of arms control for cyberspace. It gives an overview of the general architecture of arms control regimes and the complex issue of establishing and verifying compliance with agreements. Building on these considerations, the chapter presents important treaties and first approaches, including the Wassenaar Arrangement, the recommendations of the OSCE, and the UN GGE 2015.

Published as: Reinhold, T., & Reuter, C. (2019a, March 13). Arms Control and its Applicability to Cyberspace. In C. Reuter (Ed.), *Information Technology for Peace and Security: IT Applications and Infrastructures in Conflicts, Crises, War, and Peace* (pp. 207–231). Springer Vieweg. https://doi.org/10.1007/978-3-658-25652-4_10

AUTHOR STATEMENT Thomas Reinhold has written the chapters 8.1 to 8.4. Thomas Reinhold has performed the analysis and comparison of prior arms control approaches and confidence building measures and performed the analysis of already existing early approaches and regulatory measures of malicious activities in cyberspace as well as the limitations that come with this domain. Christian Reuter supported the refining of the argumentation and the structuring of the paper.

Chapter 9: "Challenges for Cyber Arms Control: A Qualitative Expert Interview Study" focuses on the debates regarding these challenges and the obstacles of dual-use, proliferation, constant technological progress, the importance of the private sector, difficulties in defining and verifying the weapon, and difficulties in attributing attacks. By employing a literature review as well as qualitative expert interviews, the chapter aims to provide a state-of-the-art perspective on these topics, the question to what extent does expert knowledge align with the challenges discussed in the literature and what conclusions can be drawn for further debates.

Accepted and getting published as: Reinhold, T., Pleil, H., & Reuter, C. (2023). Challenges for Cyber Arms Control: A Qualitative Expert Interview Study. *Zeitschrift für Außen- und Sicherheitspolitik (ZfAS)*. https://doi.org/10.1007/s12399-023-00960-w

AUTHOR STATEMENT The paper constitutes a joint work with Helene Pleil (Digital Society Institute at the European School of Management and Technology Berlin) and Christian Reuter. As corresponding and leading author, Thomas Reinhold was in charge of the overall research design and management of the paper. Following the structure of the paper, Thomas Reinhold has written the introduction (9.1), supported by Helene Pleil and developed the theoretical perspective as well as the related work (9.2) together with Helene Pleil. The methodology (9.3) has been mainly written by Helene Pleil, but has been thematically and methodically supervised and led by Thomas Reinhold who also contributed the research question, supported choosing the interview partners and the creation of

the interview guidelines. The interviews and their coding have been performed by Helene Pleil. The analysis of the interviews and the empirical findings (9.4) have been performed by Helene Pleil and Thomas Reinhold together. The discussion and especially the interpretation of the results regarding the research context of arms control (9.5) as well as the limitations, future work and the final conclusions (9.6), has mainly been written by Thomas Reinhold based on a draft version written by Helene Pleil. Christian Reuter supported especially in creating the paper structure and streamlining the arguing of our contribution and refining the paper.

1.5.2 *Part B: Threats From Malicious Activities in Cyberspace and Technological Trends*

Chapter 10: "Towards a Cyberweapons Assessment Model – Assessment of the Technical Features of Malicious Software" analyzes the current perspectives on cyberweapons, identifying their weaknesses of being either based on assumptions about adversarial actors or being applicable only after the usage of a malicious tool. In contrast to these approaches, it presents an indicator-based assessment model based on specific functional aspects of malware and parameters that can be measured prior to the application of malicious software. This enables the categorization of malicious tools as cyberweapons.

Published as: Reinhold, T., & Reuter, C. (2021). Towards a Cyber Weapons Assessment Model – Assessment of the Technical Features of Malicious Software. *IEEE Transactions on Technology and Society*, *3*(3), 226–239. https://doi.org/10.1109/TTS.2021.3131817

Author Statement Thomas Reinhold has written the chapters 10.1 to 10.5. Thomas Reinhold has performed the literature analysis of the current perspectives on cyberweapons, performed the comparison of these different approaches and developed the research question regarding a technical assessment model for this classification. Thomas Reinhold performed the assessment of which technical parameters exist that are suitable for this approach, developed the assessment model to classify malicious software as "cyberweapon" based on technical parameters and performed the analysis of this approach based on two examples cases. Based on this, Thomas Reinhold performed the assessment of the limitation of this approach and further necessary research. Christian Reuter supported in refining the argumentation, defining the research question, and in structuring as well as streamlining the paper.

Chapter 11: "Spotting the Bear: Credible Attribution and Russian Operations in Cyberspace" discusses the tools and techniques that could help to identify the hackers who have conducted a cyber-operation and elaborates why credible attribution in the case of cyberattacks carried out or masterminded by state actors is so challenging. Based on case examples regarding malicious activities by presumably Russian based actors, the paper addresses the technical, intelligence and geopolitical aspects of attribution and highlights, explaining what attribution is and highlighting its importance in the domain of cybersecurity.

Published as: Herpig, S., & Reinhold, T. (2018, October). Spotting the Bear: Credible Attribution and Russian Operations in Cyberspace. In N. Popescu & S. Secrieru (Eds.), *Hacks, leaks and disruptions: Russian cyber strategies* (pp. 33–42, Vol. 148). European Union Institute for Security Studies (EUISS). https://www.jstor.org/stable/resrep21140.7

AUTHOR STATEMENT The text constitutes a joint work with Sven Herpig (Stiftung Neue Verantwortung, SNV). As corresponding and leading author, Thomas Reinhold was in charge of the overall design and management of the paper. Following the structure of the paper, Thomas Reinhold has written the introduction that explains the attribution problem (11.1) as well as analyzed and discussed the technical aspects of attributing malicious cyber activities (11.2.1). Sven Herpig provided the subchapter with regard to the intelligence aspects of attribution (11.2.2). The discussion of the geopolitical aspects of attribution was a joint effort (11.2.3). The conclusion of the paper (11.3) has been written by Thomas Reinhold with support of Sven Herpig, especially with regard to the selection of the example cases.

Chapter 12: "Wannacry About the Tragedy of the Commons? Game-theory and the Failure of Global Vulnerability Disclosure" analyses the incentives and challenges of vulnerability disclosure by state actors and the problem of colliding interests between cybersecurity on the one hand and purpose of foreign espionage, surveillance and law enforcement on the other hand. Based on the game theory of "tragedy of the commons" and the so-called "prisoners dilemma", the article develops a set of international best practices to escape this dilemma, focussing on the questions of the smallest common denominator of such a global vulnerability disclosure regime and under what conditions such an agreement could be reached, based on the case example of the 2017 incidents regarding EternalBlue.

Published as: Schulze, M., & Reinhold, T. (2018). Wannacry About the Tragedy of the Commons? Game-Theory and the Failure of Global Vulnerability Disclosure. *European Conference on Information Warfare and Security, ECCWS*, *2018-June*, 454–463. https://www.proquest.com/openview/f6ccddd62973bd8997c3fcd40951f4f1/1?cbl=396497&pq-origsite=gscholar&parentSessionId=7jm9tc94UMKaTk9pAtTjzd%2BhJYdl8V55qGHqrUpnUM8%3D

AUTHOR STATEMENT The text constitutes a joint work with Matthias Schulze (German Institute for International and Security Affairs, SWP). As corresponding and leading author, Thomas Reinhold was in charge of the overall research design and management of the paper. Following the structure of the paper, Thomas Reinhold has written the introduction and identified the research gap (12.1). Matthias Schulze provided the theoretical concepts of game theory and specifically the prisoner's dilemma (12.2). The application of these concepts on vulnerability disclosure, digital goods and the behavior of state actors (12.3) has been carried out by Thomas Reinhold. Thomas Reinhold also developed the recommendations for inter-state cooperation (12.4) in the process of the disclosure of vulnerabilities with support of Matthias Schulze with regard to game theory peculiarities. The conclusion of this paper (12.5) was a joint effort. Matthias Schulze presented the paper to the ECCWS conference.

Chapter 13: "The Digital Divide in State Vulnerability to Submarine Communications Cable Failure" analyses the dependencies of coastal and island states on submarine communication cables (SCC) as part of the physical internet infrastructure to provide internet connectivity. The paper models the worldwide SCC network based on publicly available data, analyses the global network properties with a focus on putting the remaining bandwidth capacities in three different failure scenario simulations of SCC breakdowns. It identifies 15 highly vulnerable states and overseas territories, and another 28 territories that are classified as partially vulnerable to SCC failures and contributes to a better assessment of the necessity of preventive protection measures and redundancy of critical telecommunication infrastructures in states and territories characterized by high and medium vulnerability.

Published as: Franken, J., Reinhold, T., Reichert, L., & Reuter, C. (2022). The Digital Divide in State Vulnerability to Submarine Communications Cable Failure. *International Journal of Critical Infrastructure Protection*, *38*. https://doi.org/10.1016/j.ijcip.2022.100522

AUTHOR STATEMENT The text constitutes a joint work with Jonas Franken, Lilian Reichert, and Christian Reuter. The idea and the research question were drafted by Thomas Reinhold. Thomas Reinhold came up with the initial idea to research the global submarine ICT infrastructures with regard to its dependencies for states as well as the methodological approach of mapping and combining public data for the analysis. Jonas Franken has drafted chapters 13.1 and 13.2, where Thomas Reinhold has improved both chapters with regard to the research context and the research question. Thomas Reinhold was responsible for the development of the research method in chapter 13.3 and the data mapping design. Jonas Franken compiled two unified data sets from public data, conducted the data analysis, and drafted chapters 13.3 and 13.4, which have both been extended and refined by Thomas Reinhold. The discussion in chapter 13.5 was performed by Jonas Franken and Thomas Reinhold together and drafted by Jonas Franken, with support and refinement of Thomas Reinhold. The conclusion in chapter 13.6 was drafted by Jonas Franken and augmented by Thomas Reinhold. Lilian Reichert provided constant feedback focused on method and related works, as well as conceptual critique and thorough proofreading. Christian Reuter was a general advisor of this work and contributed with continuous feedback during all phases of the paper writing process.

Chapter 14: "Cyberweapons and Artificial Intelligence – Impact, Influence and the Challenges for Arms Control" is dedicated to the current boost of artificial intelligence algorithms and – among others – their application for tools used in cyberspace. It analyses already existing applications as well as current and future trends of this development, discusses their influence towards defensive and offensive usage of cyber tools and which specific challenges cyber arms control measures will face in view of these technological advances.

Published as: Reinhold, T., & Reuter, C. (2022, October 9). Cyber Weapons and Artificial Intelligence: Impact, Influence and the Challenges for Arms Control. In T. Reinhold & N. Schörnig (Eds.), *Armament, Arms Control and Artificial*

Intelligence: The Janus-faced Nature of Machine Learning in the Military Realm (pp. 145–158). Springer International Publishing. https://doi.org/10.1007/978-3-031-11043-6_11

Author Statement Thomas Reinhold has written the chapters 14.1 to 14.7, starting with the analysis of how cyberattacks and their defense are currently performed and, based on a literature analysis, how this could evolve. Based on this, Thomas Reinhold performed an assessment of the current trends and leap-forwards in the development of artificial intelligence applications, how this could affect the militarization of cyberspace and which problems arise from the fusion of artificial intelligence and malicious cyber activities. Considering this, Thomas Reinhold analyzed the impact of these technological trends on top of the already existing challenges of developing arms control measures for cyberspace, but also how artificial intelligence might help to develop arms control for the cyberspace domain. Thomas Reinhold provided a conclusion and outlook towards this ongoing trend. Christian Reuter supported this work throughout the process of developing the paper and the research question, helped to refine the argumentation as well as with structuring and streamlining the paper.

1.5.3 *Part C: Approaches for the Peaceful Development of Cyberspace*

Chapter 15: "Preventing the Escalation of Cyberconflicts: Towards an Approach to Plausibly Assure the Non-Involvement in a Cyberattack" discusses the ambiguity of digital data, especially regarding the attribution of cyberattacks and the problem of misinterpreting available information, which increases the risk of misinformed reactions and conflict escalation. In order to reduce this risk, this paper proposes a transparency system based on technologies which usually already exist for IT security measures that an accused actor in a specific incident can use to provide credible information which plausibly assures his non-involvement. The paper analyses the technical requirements, presents the technical concept based and discusses the necessary adjustments to existing IT networks for its implementation. With regard to its intended benefit of conflict de-escalation, the limitations of this approach are assessed, regarding the technical limits as well as the political motivation, the behavior of states and cheating.

Published as: Reinhold, T., & Reuter, C. (2023a). Preventing the Escalation of Cyber Conflicts: Towards an Approach To Plausibly Assure the Non-Involvement in a Cyberattack. *Zeitschrift für Friedens- und Konfliktforschung (ZeFKo)*. https://doi.org/10.1007/s42597-023-00099-7

Author Statement Thomas Reinhold has written the chapters 15.1 to 15.6. Thomas Reinhold analyzed the current state of cyberconflicts, illustrated by two exemplary cases, presented a schematic model of such incidents and preformed the analysis of current perspectives on de-escalating such tense situations. Thomas Reinhold performed the assessment of the technical characteristics of cyberspace that support the escalation of cyberconflicts, developed the outline for a technical system that could help to reduce this escalation risk based on an analysis of the political and technical requirements for such an approach. Regarding the

implementation of this measure, Thomas Reinhold discussed the limitations and pitfalls of the systems and presented the outlook for further research. Christian Reuter supported the refinement of the research question, the argumentation of the measures, the structuring of the paper and regarding the discussion of its limitations.

Chapter 16: "ExTRUST: Reducing Exploit Stockpiles with a Privacy-Preserving Depletion System for Inter-State Relationship" analyses the situation of state actors stockpiling vulnerabilities for strategic or law enforcement interest, thus resulting in a situation where multiple states probably stockpile at least some identical exploits. This paper proposes a privacy-preserving approach that allows multiple state parties to privately compare exploits and check for items that occur in multiple stockpiles without disclosing them, thus identifying matches known by multiple parties that can potentially be published to be fixed. Although ExTRUST focuses on the context of inter-state relations, and considers the special constraints of cooperation between state actors and the requirements within such environments, it can also be used for other zero-trust use cases, such as bug-bounty programs and the confidential check if a reported vulnerability is a zero-day exploit or already known.

Published as: Reinhold, T., Kühn, P., Günther, D., Schneider, T., & Reuter, C. (2023). EXTRUST: Reducing Exploit Stockpiles With a Privacy-Preserving Depletion System for Inter-State Relationship. *IEEE Transactions on Technology and Society*. https://doi.org/10.1109/TTS.2023.3280356

Author Statement The text constitutes a joint work of Thomas Reinhold with Philipp Kühn, Christian Reuter as well as Daniel Günther and Thomas Schneider (both part of the research group Cryptography and Privacy Engineering at the Department of Computer Science at the Technical University of Darmstadt). As corresponding and leading author, Thomas Reinhold was in charge of the overall research design and management of the paper. Following the structure of the paper, Thomas Reinhold has written the introduction (16.1). The related work chapter (16.2) was a joint work of Philipp Kühn, Daniel Günther and Thomas Reinhold, who analyzed and identified the research gap (16.2.4). Thomas Reinhold carried out the analysis of the requirements (16.3). Philipp Kühn provided the model for an Identifier of Vulnerabilities (16.4) and developed the Blockchain-based prototype under the guidance of Thomas Reinhold (Ch. 16.5.1 and 16.5.2). The evaluation of this approach (16.5.3) is a collaborative work of Philipp Kühn – who provided the analysis of the identifiers – and Thomas Reinhold, who also carried out the technical and conceptual evaluation. The results and findings had been used for the ExTRUST system using Multi-Party Computation (16.6). Its development and evaluation had been carried out by Daniel Günther and Thomas Schneider. The discussion of our work was a joint effort with Daniel Günther, who conducted the Multi-Party Computation related evaluation (Ch. 16.7.1 and 16.7.2). Thomas Reinhold has been responsible for the conceptual evaluation of this approach with regard to the application context of arms control (Ch. 16.7.1 and 16.7.2), further application scenarios (16.7.3) as well as the conclusion and future work (Ch. 16.8). Christian Reuter and Thomas Schneider supported with discussing the related work, refining the paper structure and arguing our contribution, as well as with support in the revision processes.

Chapter 17: "Verification in Cyberspace" analyses one of the pillars of arms control and non-proliferation treaties, that provides practical measures to mutually control the treaty compliance, and the challenges regarding its development for cyberspace due to its unique technical characteristics. Based on these peculiarities and the fact that cyberspace is a man-made domain, it discusses the possibilities for adjusting its technical settings, rules and principles and how this may help to reduce the threat of ongoing militarization. Offering some alternatives, the chapter elaborates on suitable and measurable parameters for this domain and presents potentially useful verification approaches.

Published as: Reinhold, T., & Reuter, C. (2019c, March 13). Verification in Cyberspace. In C. Reuter (Ed.), *Information Technology for Peace and Security: IT Applications and Infrastructures in Conflicts, Crises, War, and Peace* (pp. 257–275). Springer Vieweg. https://doi.org/10.1007/978-3-658-25652-4_12

AUTHOR STATEMENT Thomas Reinhold has written the chapters 17.1 to 17.5. Thomas Reinhold has performed the analysis regarding the research question, how verification (in the context of arms control) can work for the cyberspace domain and which impediments exist towards the application of established approaches. Based on these, Thomas Reinhold has performed the assessment of which approaches exist that are based on the technical features of cyberspace, discussed how they can be implemented and the limitations of this approach. Christian Reuter supported the refinement of the argumentation and the structuring of the paper.

2

FINDINGS PART A: CONCEPTS AND CHALLENGES OF PEACE IN CYBERSPACE

The goal of part A is to identify the domain-specific characteristics of cyberspace that make it difficult to transfer established arms control procedures to cyberspace, in order to reflect the current state of the debates and perspectives and to identify the limits and specific boundary conditions for the development of cyber arms control and de-escalation of cyber conflicts. This includes analyzing the behavior of states in this domain and the question of what constitutes a cyberwar.

2.1 Malicious Actors in Cyberspace

Although cyberspace has been a domain of diverse actors – from activists, intelligence services and criminal groups – for at least two decades, the revelation of Stuxnet in 2010 had a major impact on this situation. The paper "From Cyberwar to Cyberpeace" (Chapter 6) analyses the development of the militarization of cyberspace over the last decade based on a literature review and an assessment of international security policies. It shows that Stuxnet, even if not the first presumably state-led cyberattack, demonstrated two major aspects with its complex attack, the enormous amount of preparation and commitment as well as its severe and highly vulnerable target of a nuclear enrichment facility. First, at least some state actors have the knowledge, intention and resources to develop tailor-made malicious software and perform years-long hidden attacks on IT systems. Secondly, even critical infrastructures, which had been assumed to have special safety and protection measures in place due to their criticality, are vulnerable to cyberattacks and thus, that societies are vulnerable due to the dependencies on these systems. This insight had a major impact on the security politics of many states, as governments have been increasingly perceiving cyberspace as a military domain. According to an extensive study of national security policies held by the United Nations Institute for Disarmament Research in 2013, at least 47 states operated military cyber programs, of which ten nations had a nominally offensive intention (UNIDIR, 2013). Although since then no other comparable comprehensive study has been performed, Wikipedia currently lists 66 states that declare having military cyber programs (Wikipedia, 2023), whereas the "Cyber operations' tracker" database of the Council for Foreign Relations currently lists 47 states that "are suspected of [have been] sponsoring cyber operations" since 2005 (Council on Foreign Relations, 2023). Even NATO has integrated defense in cyberspace into collective defense according to article 5 at its Warsaw Summit 2016 and the EU has established a common cyber defense strategy (EU-Commission, 2022).

Besides the sheer amount of newly established military cyber forces and their activities, the paper finds that the strategic approaches behind military cyber capabilities also

changed. For one, defending against threats in cyberspace proves to be a challenge, and the defense perimeter moves from the defenders' own IT systems into the systems of current or presumed malicious actors. This trend of "active defense" or "forward defense" (Herpig et al., 2020) builds upon the rational that creating friction in the systems of an attacker might support mitigating an ongoing threat and denying the attacker's own cyberattack tool-set and infrastructure. Some countries extend this to being active in foreign IT systems even in peacetime – an approach called "persistent threat" – in order to completely prevent attempted attacks and disrupt opportunities. On the other hand, the comparably cheap and technological off-the-shelf availability of cyberattack knowledge and tools led to an "empowerment" of small states. For instance, North Korea, which would not be able to establish itself as relevant players in the field of conventional armament due to financial and technical resources, has become a powerful and active player in cyberspace, thus establishing itself as a regional power.

2.2 The Nature of Cyberwarfare

One of the main questions regarding the militarization of cyberspace was what a military conflict in or with the support of the cyber domain will look like and which range of cyberspace activities have to be expected.

The study "Military Cyber Activities in Russia's War Against Ukraine and Their Significance for the Debates on the Containment of a Cyberwar" (Chapter 7) focused on this question, based on a structured literature analysis and the comparison of prevailing theories of an anticipated state-led cyberwar. The paper finds that experts expected a continuum between a primarily intelligence-focused meaning of cyberweapons on the one hand and the use of cyberweapons as a replacement for conventional warfare and boots on the ground on the other. In theories of the intelligence-focused meaning of cyber means of action, the domain of cyberspace serves primarily for intelligence and military information gathering and, in open military conflicts, for tactical and strategic planning and command and control. Accordingly, such a form of cyberwarfare is characterized primarily by extensive but cautious and covert cyber operations that aim at information gathering and do not pursue damaging intentions. On the other end, the "boots on the ground replacement" perspective expects cyber activities that cause severe damages, including the loss of life. Such an approach presupposes comprehensive planning and preparation in peacetime as part of military strategic planning to identify relevant military IT targets in cyberspace, infiltrate them and install, hide and maintain backdoors in order to have the option for deploying the payload in times of conflict. This perspective includes the concern that so-called critical infrastructures might be of particular interest for potential cyberattacks due to the potentially disruptive effects and large-scale impairments to a state that can directly limit the military options for action.

Russia's invasion of Ukraine and the ensuing war have, among many other security certainties, demonstrated for the first time the role of cyberspace in an open war. In contrast to the analyzed expectations, a review of publically available media and other coverages of the events since the start of the war in 2022 shows that a comprehensive cyberwar in which military action in cyberspace plays a decisive role has so far not occurred. This is probably a result of Russia's specific situation, tactical planning failures and misjudgments, as well as the immediate, very strong support of Ukraine,

especially in cyberspace. An analysis of scientific and political research papers in combination with critical inclusion of publically available news and OSINT[1] information showed, that apart from a prelude phase of the war that included concerted cyberattacks, conventional military force has been the weapon of choice for Russia even during the mid-2022 initiated phase of attacks specifically on critical infrastructures. This suggests that a full cyberwar as a boots-on-the-ground substitute is rather unlikely. Nevertheless, cyberspace will probably continue to gain in importance as a military domain, especially with regard to tactical planning in the area of interdiction and disruption of communications and supply infrastructures, which we have seen in Russia's war on a large scale. This development can also result in even more cyber activities of state-led actors in foreign IT systems in peacetime for intelligence gathering and either preparations for armed conflicts or precautions. A specific development that had not been expected and discussed in prior theories concerns the role of non-state actors, which to a certain degree have been motivated and in some parts allegedly ordered by state institutions, that have undertaken enormous and unexpected activities in cyberspace on both sides. Their activities at least contributed to undermining Russia's restrictive information policy regarding the war, created a lot of friction on both sides of the conflict and bound IT security resources. This development raises concerns about how actions of non-state actors can be contained in future conflict situations in order to be able to control conflict dynamics, prevent unintended conflict escalations, protect the civilian population and avoid an international proliferation of the actors involved in the war. Finally, it opens the question of the extent to which civilian actors could become military combatants and what the consequences would be for the "ius in bello" (the law in times of war) in particular and the perspective on the war in general if cyber-combatants presumed to be actively intervening in a conflict are completely decoupled from the territorial limitations of a belligerent confrontation.

2.3 Arms Control and Its Challenges in Cyberspace

Regarding the analyzed challenges of militarization, the question arises of how a peaceful development of this domain and a reduction of the current threats can be achieved.

The paper "Arms Control and its Applicability to Cyberspace" (Chapter 8) analyzes, based on a literature review, how prior arms control approaches can be applied in cyberspace, which challenges exist and where new solutions that take the specifics of cyberspace into account are necessary. It summarizes the development and implementation of different arms control measures, that can generally be divided into three different parts: war prevention and the reduction of conflict probability, damage limitation in the event of armed conflicts and reduction of armament-related costs. It assesses that, besides ongoing political processes of finding a common understanding and terminology of this new technology, the domain cyberspace itself has some very specific characteristics that are different from the other domains land, air and sea. The paper identifies and analyzes these characteristics, that pose a problem for transferring previous arms control approaches. This includes the virtuality of this domain, the non-physical representation of code, the seamless duplication of data, the strong dual-use aspect of cyber tools and the problematic differentiation between defensive and offensive usage as well as the

[1] Open Source Intelligence

complex and complicated assignment of responsibilities for malicious activities in this domain.

The study "Challenges for Cyber Arms Control: A Qualitative Expert Interview Study" (Chapter 9) focus on the current political process under the head of the UN and the OSCE. The paper analyzes the work in the context of the "UN Group of Governmental Experts on Advancing responsible State behavior in cyberspace in the context of international security" regarding the recognition of applicable and existing rules of international law for cyberspace, the UNODA organized "Open ended working group on security of and in the use of information and communications technologies" discussions and proposals of capacity building and developing a common political ground for cyberspace debates as well as the concepts for confidence and trust building measures proposed by the OSCE. The study finds, based on a qualitative interview study, that a lot of ground has to get covered and highlights, that – according to experts – essential problems still exists. This especially concerns the lack of a binding definition for the term cyberweapon, the dual-use dilemma and the lack of a working distinction between civil and military IT products, as well as the still insufficient multi-stakeholder inclusion of the private sector. The study further highlighted missing solutions like for the yet unsolved challenge of verification measures for cyberspace or a reliable attribution mechanism, which involves technical as well as political dimensions to attribute malicious cyber activities to their origin.

Finally, a finding that impacts all the above-mentioned challenges and processes, is that states are currently still in the discovery phase as to what advantages and opportunities cyberspace could offer them. As a result, they are just beginning to perceive cyber tools as strategically valuable and have diverging interests, even between states that otherwise share common values and goals. The many possible use cases of cyber instruments are considered highly relevant and worthwhile for states, as they do not want to forego the associated advantages, and especially espionage activities tend to become unexpected norm setters for state behavior in this domain. This results in a situation where malicious cyber tools are being widely used by various states, while any attempt of an agreement to limit this behavior encounters the states' fear of disadvantages vis-à-vis competitors. In addition, states have to deal with a variety of non-state actors that also posses such tools. These actors are therefore relevant for the restriction of the use of cyberweapons, but are probably not addressable by arms control agreements that are usually concluded exclusively between states. A situation which is yet completely unsolved.

3

FINDINGS PART B: THREATS FROM MALICIOUS ACTIVITIES IN CYBERSPACE AND TECHNOLOGICAL TRENDS

The publications of part B analyze the challenges that arise from the militarization of cyberspace as well as current technological trends that undermine the security of this domain in order to identify the specific threats to be addressed with arms control and to identify starting points for the development of practical measures.

3.1 Assessment Model for Cyberweapons

Among the findings from part A, there was a distinct conclusion that cyberspace has some unique features that strongly differ from other domains like air, land and sea. This situation results in some special threats and challenges, especially for any approaches to a de-escalation of state-led cyberconflicts and cyber arms control. Part B of the publication of this doctoral thesis focuses on these challenges, analyses the underlying problems and their influence on the development of cyber arms control measures.

One of the primary challenges regarding the regulation of malicious activities in cyberspace is the missing of an internationally binding, commonly agreed upon definition of what constitutes a cyberweapon, a question that has been researched in the paper "Towards a Cyberweapons Assessment Model – Assessment of the Technical Features of Malicious Software" (Chapter 10). Via a structured analytical review of existing approaches that focus on this classification of malicious cyber tools, the study reveals that these utilize either an assessment of the intent of the attacker or the purposed potential or actual triggered effects of the assessed tool. While these are valuable approaches for political processes and debates, that focus primarily on norms like the dos and don'ts for state behavior in cyberspace, they lack a specific feature. These approaches are not applicable in advance of a specific incident and their presumed intent will always be influenced by speculation, political and strategic considerations, and the interests of various relevant actors. In contrast, arms control, as well as any regulation towards the non-use of specific tools as a preventive approach, has a necessity for assessment measures that are applicable regardless of subjective considerations and especially before the tool has been used. Given this requirement, the paper developed an assessment model that is based on existing technical parameters of IT soft- and hardware. The assessment model deconstructs potential military weapon systems into their components and the multitude of different interoperating parts, materials and underlying technologies for their development and production. The model identifies several parameters and indicators that can be collected, tracked, or counted, provides evaluation ranges for each parameter and proposes an assessment schema that reflects the different steps from development to deployment of potential weapons. The evaluation of three selected example cases showed that the classification of potential cyberweapons is possible based

on distinguishable, measurable features of a malicious cyber tool regardless of its prior usage and independently of speculations about its intent. Although the assessment model does not claim to be exhaustive, nor can such a generalization that considers a broad range of parameters deliver "sharp edges", it nevertheless provides a standardized and unified procedure to determine if a specific malware can be considered a cyberweapon. It is therefore meant as a decision support for case-by-case assessments, which exactly fits the necessity of cyber arms control.

3.2 The Attribution Problem

A second, most pressing problem that is directly rooted within the technical features of cyberspace, like its virtuality and the multiple technical solutions to hide the path of an attacker throughout the internet, is the so-called attribution problem. This term describes the challenge of tracking a cyberattack to its origin, in order to either politically confront the attacker or to initiate countermeasures to mitigate an imminent threat. Being able to pinpoint the source of an attack is also an essential requirement of international humanitarian law and especially the UN Charta which allows measures of self-defense for a threatened state as an exception to the otherwise complete prohibition of state violence. The study "Spotting the Bear: Credible Attribution and Russian Operations in Cyberspace" (Chapter 11) focuses on this problem based on examples of malicious cyber activities from Russian actors by analyzing previous cases of cyberattacks and their attribution. It revealed that attribution could have two further audiences to get addressed and aims to reach, besides the perspective on the attacker itself. It showed that on the one hand, the so-called public attribution of an attack refers to convincing the public of a state of a certain assessment in order to allow the government to choose from a wider range of policy options. On the other hand, the so-called transnational attribution refers to convincing allies, bilaterally or as a whole (for example through NATO), of the validity of the attribution.

Regarding the attribution process and the determination of the question "who is the attacker", the case analysis finds that attribution usually has three dimensions that have to be taken into account. First, the technical dimension of gathering information about the network connections and activities of an attacker prior to an attack or whilst in progress. In order to support this data collection, appropriate measures must exist in the domestic IT systems and networks of a state ("inner scope measures") and in foreign IT systems ("outer scope measures") that an attacker has presumably used. The technical tools can be further distinguished into preventive measures, that constantly observe, collect and store data, and reactive measures that can be used to mark and track down an attacker during an ongoing operation. The analysis showed that this dimension is not sufficient, either because the required data is not completely available, but also because the physical origin of an attacking IT system cannot automatically be equated with the origin of the attacker. The second dimension that needs to be taken into account, in order to draw the right conclusions from the gathered information, regards the intelligence aspect of attribution. This relates especially to human and signal intelligence, either gathered through a state's own means or accessed via shared resources by allies. Such information can help to filter or assess the technical information and connect it to previous or other current events, as well as databases of known tools or tactics. The downside of this dimension is whether collected proof can be presented

to international organizations (e.g. UN, NATO) and/or the public. This would likely mean exposure of the intelligence operation, possibly decrease the likelihood of it still being effective in the future, and even provide attackers with information to counter-respond or circumvent being tracked. This underlines what our analysis also showed, that attribution needs a third dimension, the geopolitical assessment, which supplements gathering hard facts and concrete evidence in order to validate the assessment. This geopolitical perspective ultimately focuses on the attacker's motivation and the two questions who would directly and most significantly benefit from the attack or from the attack being mistakenly attributed to a third party and to what extent the attack could be part of a larger deception strategy of a state.

The analysis demonstrated that attribution and accusations in specific conflicts need to be based on a credible, evidence-based argumentation that has to be made by the affected state. Enabling states to conduct more severe responses to cyberattacks would require public and international attribution. This in turn leads to a Catch-22 situation in intelligence sharing. Pointing the finger at the usual suspect without credible attribution when a cyberattack occurs not only further emboldens the attacker in their projection of power (while they continue to deny responsibility) but fails to sufficiently convince the public or the international community.

3.3 A State's Dilemma with Vulnerabilities

Vulnerabilities in IT software and hardware products are the essential material of most malicious cyber activities. Intelligence services, law enforcement agencies or military forces being active in cyberspace and performing their duties, usually rely on vulnerabilities for a specific IT system or product that can be turned into an exploit to gather access to computers or networks. Therefore, these organizations have an inherent interest in withholding collected vulnerabilities and stockpile them, in order to have them available when needed. This creates a dilemma, as any potent vulnerability that has not been disclosed and therefore has not been patched remains open for exploitation. This potentially leaves a lot of IT systems vulnerable, also threatening civil or commercial IT systems. Moreover, states further risk that another actor finds the very same vulnerability and exploits it. Thus, states engaging in cyber-operations have conflicting interests: stockpiling vulnerabilities for a national advantage or disclosing vulnerabilities to increase cybersecurity for everyone.

The paper "Wannacry About the Tragedy of the Commons? Game-theory and the Failure of Global Vulnerability Disclosure" (Chapter 12) assesses this dilemma based on a case study analysis of the WannaCry malware incident from 2017. It develops a game-theory-driven model of the conflicting interests of states in order to derive preliminary strategies to achieve cooperation in disclosing vulnerabilities on a global scale, using theories of international relations. The used so-called "prisoner's dilemma" presents a thought model of a situation of uncertainty which applies to the strategic rationality behind hoarding zero-day exploits that has three possible outcomes: (a) both actors cooperate and responsibly disclose vulnerabilities, gaining an increased level of cybersecurity as a common shared good; (b) only one of the actors discloses his vulnerabilities, whilst the other actor keeps his vulnerabilities, leaving him at a strategic advantage; and (c) a situation where both actors keep their vulnerabilities hidden, thus being and staying

vulnerable to cyberattacks. A core insight from the model is that any kind of uncertainty in combination with missing communication leads to the worst possible variants, as actors are incentivized to not give up their strategic possibilities.

An easy-to-reach incentive for cooperation can be mutual agreements about threat and vulnerability sharing between partners, which resembles a reiterated or tit-for-tat cooperation. The analysis showed that such sharing platforms hugely benefit those participating in them. They further could be utilized to raise the costs of non-cooperation by the exclusiveness of time-critical threat warnings, where non-partners will get informed only with the official patch release and the vulnerability warning of the manufacturers. A different analyzed approach is to lower the benefits of non-cooperation by shortening the life cycle and thus the utility of vulnerabilities via legally binding definitions and procedures on how long an actor could keep the knowledge of vulnerabilities until it needs to get disclosed, a so-called "vulnerability equity process" (VEP). The utilization of vulnerabilities can also be reduced by further fostering or mandatorily regulating internationally standardized bug-bounty programs for all software vendors, which can be economically feasible. More radical suggestions of the study include changing the pay-off structure for vulnerability brokers and cyber-criminals by drying out the vulnerability black market. This could, for example, be reached by an international vulnerability purchase program, in which states and corporations would share the financial burden in buying all available vulnerabilities from the market and encourage ethical hackers to disclose them for more competitive rewards than typical bug bounty programs. A minimum rule that has actually been followed in the ShadowBroker incident – the incident that revealed the origins of the later-on performed WannaCry attacks – is the obligation of states to inform IT product vendors if a nation's cyber-stockpile gets stolen. This approach can even be institutionalized into an international organization, much like the WHO or the IAEA, that could standardize, collect and proliferate notifications of cyber incidents.

3.4 Threats and Vulnerabilities of the Global Internet Backbone

A special threat of malicious activities in cyberspace that has been emphasized during Russia's war against Ukraine is, that the global IT infrastructures of the internet are becoming a target themselves in order to undermine civil or military communication capabilities or to create fear and uncertainty. Additionally, the majority of the global market of Internet backbone vendors is owned by private companies that either can become the plaything of global national strategies or – like in the case of China – are under the direct influence of a state. Given the dependencies on these infrastructures, which are the focus of the study "The Digital Divide in State Vulnerability to Submarine Communications Cable Failure" (Chapter 13), the global submarine fiber optic communication cables (SCC) as the backbone of the whole internet are of the most critical reliance for most parts of the world, modern societies and economies. More than 98% of international online communication is handled via fiber optic cables laid in the world's oceans. While there is a strong global east-west SCC dominance, the situation highly differs from a global north-south perspective or regions with many islands. Even states that are mostly surrounded by terrestrial borders or that have a strong land-based supply with internet and communication capacities rely on submarine cables for communication with their embassies due to economic connections or for military forces abroad. In

contrast to this dependency, over 80% of the 1.3 million kilometers of active submarine fiber optic cables are currently located in an inaccessible deep sea below 1500 m depth, which makes it impossible for authorities or private companies to ensure continuous surveillance and physical protection of them.

The paper analyses this unbalanced situation and aims to quantify the dependency of states on SCC and to identify vulnerabilities. In order to realize this approach, the paper uses a network modeling and analysis of the global SCC network, in addition with publically available data of the cables' relevant factors like the life span of each cable, network bandwidth and redundancies. It further includes a socio-economic development factor for each state as a measure of economic strength to build cables, as well as the dependency on the internet connectivity of the society and economy. Based on this information, a network of nodes has been created (representing states and network peering points) and edges, representing the submarine cables. An analysis of this model via quantitative network analysis revealed, that for the vast majority (n = 126) of the territories examined, there is only a low probability of an internet outage after SCC failure by natural causes like earthquakes, shark attacks etc. Although, a concerted intentional attack that would destroy more than one cable, thus reducing the redundancy, could alter this picture significantly. Besides this conclusion, 43 territories with an increased risk of cable failure were identified, 15 of which did not even have one sufficient SCC as redundancy. In addition, the study found a positive correlation between a lower redundancy level and a low socio-economic development status (developing country or least developed country). Therefore, states and territories in the Global South are more likely to be highly vulnerable to SCC faults. At the same time, they often do not offer economic incentives to implement additional SCCs. This situation might improve with the upcoming impact of emerging internet-providing technologies like low-earth-orbit satellite internet and their adoption in contexts that the study rated as vulnerable, which could provide sufficient broadband connectivity in some time without requiring the construction of fiber optic cables. From the perspective of developing countries without any backbone connection or with low redundancy levels, a crucial question will also be whether the pricing of these services will lead to a further intensification of the Global Digital Divide.

3.5 How Artificial Intelligence Will Change Malicious Cyber Tools

Whereas the weaponization of cyber tools has been under discussion for quite some time, either from the perspective of necessary and appropriate defensive measures or as a new category for offensive planning, the emergence of improved algorithms in artificial intelligence and machine learning (AI/ML) and its recent technological leap is probably going to changes this field of application drastically. This situation is analyzed in the study "Cyberweapons and Artificial Intelligence – Impact, Influence and the Challenges for Arms Control" (Chapter 14). Based on a comparison of artificial intelligence and machine learning with cyber tools, the study highlights that – from a technological perspective – both are natural siblings. They are based on complex computer code that is developed and deployed within the same domain, require to a considerable extent similar know-how in programming, code logic and software life cycle management and are usually based on modularized, extendable and interchangeable software frameworks to provide adaptability and expandability. Additionally, computer code offers optimal

conditions for creating and facilitating training and testing environments for AI/ML applications, as the environment can be defined and shaped in every specific detail and according to the intended requirements. An additional literature analysis of the current technological challenges for military cyber tools revealed the problem of a growing amount of information that needs to be processed in contrast to the decreasing time to react to incidents. AI/ML algorithms and especially modern approaches such as deep learning were developed specifically for such cases, involving the processing of large amounts of data, detecting patterns and filtering out relevant information from digital noise.

Based on this assessment, the study found that using AL/ML for weapon systems introduces some essential problems on different levels. First and foremost the question of autonomy and meaningful human control, which has been discussed extensively in the context of drones and other lethal or non-lethal autonomous weapon systems (LAWS). Furthermore, the black-box character of AI raises the challenge, which measures exist that allow an understanding of the relation of the input-output processing of an AI. The study found that a dedicated field of research (XAI – Explainable Artificial Intelligence) is working on such concepts that provide tools to follow the decisions during the reasoning process (ad-hoc XAI) or the decisions to be recapped once they are made (post-hoc XAI). So far, these approaches are mere theoretical concepts that lack general applicability or are hindered by specific technical features of machine learning, such as the distributed and numerical representation of learned information. As an additional challenge, AI/ML algorithms are trained for specific situations and decisions before they are integrated into production systems. This results in a situation where operators of the finished application might be unlikely to know the specific details of the training data, nor have any chance to see, perceive or understand the assumptions and preconditions of this data. The analysis found that these aspects, in addition to the speed of reaction, possibly prevent any opportunity for double-checking decisions by human operators.

The study further found, that applying AI/ML measures especially to automatic cyber defensive operations will probably further aggravate the ambiguity of the attribution problem. Wrongful automatic conclusions about the origin of an attack could be triggered either by incorrect or insufficiently trained algorithms, biased input information or by following intentionally created false trails. In addition, an attacker with knowledge about the applied AI can possibly replicate it to use this knowledge in order to tailor their attacks either to avoid detection or to purposefully generate incorrect conclusions.

4

FINDINGS PART C: APPROACHES FOR THE PEACEFUL DEVELOPMENT OF CYBERSPACE

Part C concludes the results from the prior parts by developing practical measures for arms-control and de-escalation measures in cyberspace. It discusses how existing procedures from other areas of computer science can be adopted and applied for such measures, identifies suitable technical parameters and discusses the technical requirements and limitations for their implementation?

4.1 De-escalating Cyberconflicts

The ambiguity of digital data, that has already been discussed as a fundamental challenge regarding the attribution problem, contains the inherent risk of misinterpreting available information and therefore increases the threat of misinformed reactions. With the complex and often time-consuming process of the secure identification of the sources of malicious cyber activities declared to be impractical by many experts, especially situations with the necessity of immediate responses to counter and mitigate cyber threats could lead to misguided responses and create a momentum for the escalation of conflicts, rendering politically tense situations especially vulnerable for false flag operations. Regarding this situation, the study "Preventing the Escalation of Cyberconflicts: Towards an Approach to Plausibly Assure the Non-Involvement in a Cyberattack" (Chapter 15) assessed possible technical measures, that enable state actors to provide verifiable data that support a plausible argumentation for a non-involvement of the state's military forces and intelligence services in a specific previous or ongoing cyber incident. The analysis showed that such measures need to fulfill four key characteristics: (a) it has to cover a time frame that is long enough to satisfy an accusing actor, (b) needs to collect data from all relevant IT networks without hindering their functionality, (c) must contain details on the endpoints of all connections that had been established and should rely on information that is always accessible during the network-based data transmission and (d) has to use a tamper-proof, non-circumstantial data acquisition and storage measure that is considered trustworthy even in non-cooperative actor relationships.

The study found that the necessary information analysis and storage for such a measure usually already takes place in most cases, as an IT security measure to monitor connections, identify malicious activities, detect or track hacking attacks and for access control. Based on this conclusion, the paper developed and presents a concept for the implementation of a system for the plausible assurance of non-involvement in a cyber incident, based on already existing IT networking infrastructure and without requiring changes to an existing IT network, apart from immutable storage. Additionally, the paper found that existing techniques to anonymize logged information provide the necessary

secrecy for state actors as an incentive to implement the measure, without losing the evidential value of the collected information.

Although the radius of a possible implementation is and needs to be limited to specific networks of military forces or intelligence services to prevent the establishment of a surveillance system, this is not considered problematic regarding the research focus. Arms control measures directly aim for implementation by government institutions, which already have a special role that is usually associated with high responsibility and legal obligations. In contrast, the analysis also found that participating actors have different possibilities for dishonest behavior by using proxies or filtering the logged information. Apart from this, states and their institutions often have divergent, sometimes contradictory interests and political decisions and intentions sometimes contradict concluded agreements. Nevertheless, the study found that the main motivation for a state to establish measures that can plausibly assure their non-involvement in a specific cyber threat is their self-interest in preventing the escalation of a conflict or of being falsely held accountable for malicious activities. The de-escalation effect of the measure is directly linked to a states' credibility. In conclusion, this means that it is in a state's interest to comply with the measure in order to keep a tool at hand that would provide information for the chance of de-escalating uninvolved cyberconflicts.

4.2 Disarmament for Cyberspace by Reducing the Vulnerability Stockpiles

The findings regarding the dilemma for state agencies of using and withholding vulnerabilities already pointed out that, on the one hand, these agencies can have a justified interest in using vulnerabilities for their tasks. On the other hand, there is a lack of measures to reduce these stockpiles. Even the discussed VEP[1] is, so far, mainly a concept that has not yet been considered by many states. A major obstacle to such an approach is the reluctance of participating parties to disclose sensitive information about their own capabilities, which is generally seen as giving up tactical advantages, effectively resulting in an international arms race for offensive cyber capabilities. This leads to a situation where multiple states are likely to stockpile at least some identical exploits. Based on the rational choice consideration that identical vulnerabilities or exploits can be considered a candidate for disclosure, the paper "ExTRUST: Reducing Exploit Stockpiles with a Privacy-Preserving Depletion System for Inter-State Relationship" (Chapter 16) proposes a technical solution called ExTRUST. It is based on a multi-party computation approach that allows multiple actors to compare stockpiles for matching entries, while completely preserving their confidentiality. To enable this comparison, the paper developed and presents an approach for the unique, unambiguously machine-readable identification of vulnerabilities, based on a review of existing identification and classification measures. The approach provides a trade-off in description detail, that avoids both different vulnerabilities being described by the same identifier as well as the same vulnerability being described with different identifiers.

ExTRUST is developed for a zero-trust environment and does not rely on any preconditions of trust in advance or assumptions of good nature in order to reflect the special constraints and impediments of interacting states. The theoretical analysis, as well as the

[1] Vulnerability equity process, see chapter 12

implementation and testing, found that ExTRUST is scalable and allows parties to join or exit the measure at any time. It is decentralized and does not require a neutral authority to operate or maintain the system. The measure can be performed independently by each participating party, while feedback of a match is only available to these parties who submitted the specific obfuscated vulnerability information. The analysis found that ExTRUST is, although not real-time calculable, practical and performant for an arms control context. ExTRUST furthermore ensures the confidentiality of vulnerability or exploit information against any party, submitted data cannot be modified or corrupted by any party, the system prevents false positive intersection results and can withstand several attack scenarios like brute-force attacks, or dishonest participants.

Based on ExTRUST, states would be able to secretly check and compare their stockpiles of vulnerabilities as a practical measure of disarmament for cyberspace, supporting the reduction of the amount of undisclosed IT vulnerabilities.

4.3 Verification in Cyberspace

One of the pillars of arms control treaties is the so-called verification. It defines measures that enable treaty members to mutually control the treaty compliance by observing, counting or monitoring specific actions and their accordance with the respective rules. The amount and degree of measuring and controlling of technical parameters as part of verification are usually part of the negotiation process of an arms control treaty and are written down in the contract documents. The already discussed specific features of cyberspace, that differ from other domains, make it hard to apply established verification measures to this domain. Based on a technical analysis of these peculiarities and a literature analysis, the paper "Verification in cyberspace" (Chapter 17) compares the cyberspace challenges with selected established verification measures for nuclear, biological and chemical weapons technology.

Previous findings regarding the characterization of malicious tools as cyberweapons (See chapter 10) have already identified measurable parameters of IT systems, networks and infrastructures. Based on these findings, the paper found that, on the one hand, a set of parameters exist which are physically obvious, hard to disguise or manipulate and visible for surveillance, qualifying them for monitoring a status quo, detecting technological developments of facilities as well as revealing significant changes of established capacities. These parameters include the power supply of a facility, the available network bandwidth, the amount of peering to other networks, the cooling systems as well as the required maintenance staff. As the degree of invasiveness of each verification measure directly influences a state's incentive to participate in a treaty, a stronger monitoring that requires a deeper integration into a facility is harder to negotiate. The study found that among these parameters, the CPU and the network processing power of a facility, measures to monitor the network connections or even using technologies like Deep Packet Inspection (DPI) can reveal irregularities or discrepancies from the day-to-day activities of a facility.

Regarding the possible implementation of such measures, that paper found that the dual-use character of cyberspace can be an advantage for implementing cyberspace verification into military or intelligence service networks. Because monitoring networks

and data connections is also a core task of IT security measures, a lot of technological measures of IT facility surveillance have already been established and the results of these monitoring measures can be used and interpreted for arms control, as far as the validity of the logged information and its tamper-proof storage is given. Although negotiating cyberspace verification measures is still a challenge to be undertaken, and the overall international political situation currently does not favor arms control initiatives, this leaves room for optimism for the establishment of verification measures, which had been an important stabilization factor in the past.

5

DISCUSSION AND CONCLUSIONS

5.1 Discussion

The aim of this thesis was to identify the challenges for implementing arms control and de-escalation measures of state-led conflicts in cyberspace, to find starting points for such measures and, based on that, to exemplarily develop such measures as an impulse for further research. This thesis answers this research questions, by identifying and discussing in a first step the specific characteristics of cyberspace, of technological trends that shape this domain and of the intents, goals and behavior of the state-actors behind malicious activities. The findings show that for an increasing amount of states, cyberspace is becoming a domain for intelligence gathering or another measure to exert military force, as questioned by Buchanan and Cunningham (2020), demonstrating a long ranging strategic involvement in cyberspace to identify, establish and sustain access options already in times of peace. Especially considering the different theoretical perspectives of how cyber activities will be used in open military conflicts, as raised by Rovner (2020) or Bigelow (2019), Russia's war against Ukraine demonstrated this situation for the first time, with focused attacks on critical ICT infrastructures. The analysis of these incidents shed a light on the so-far theoretical perspective on a potential cyberwar, confirming two developments that can undermine IT security on a global scale: First, such measures are rooted in activities in foreign IT systems that can trigger unwanted effects even in times of peace and second, it advances the collection and non-disclosure of vulnerabilities. At the same time, the findings show that due to the identified technical specifics of cyberspace, the risk of miscalculations regarding the origin of an attack, misperceptions of threats and misreactions of states are high. Regarding the transfer of established arms control measures, as proposed by Hansel et al. (2018), Kühn (2014), and Maybaum and Tölle (2016), the findings highlighted big differences of cyberspace in comparison to other domains, strongly limiting this approach. On the other hand, they also revealed that a lot could still be learned from the political struggles of the last decades, regarding the underlying concepts and challenges of arms control, the political incentives as well as regulatory problems like the dual-use character, as proposed by Riebe and Reuter (2019a).

Regarding the question of which challenges arise from malicious state-led activities in cyberspace, the thesis finds the still lacking common understanding of the to-get-regulated item to be one of the main obstacles towards potential cyber arms control measures. The thesis therefore developed and proposes a classification model and appropriate assessment schema that can be used to determine the "weapon character" of a specific software, answering a question raised for example by Biller and Schmitt (2019) and thus providing a key tool for any regulation approach. In contrast to existing approaches of CCDCOE (2017) and Herr (2014) or Dewar (2017) and Orye and Maennel (2019), this measure is applicable prior to the actual usage of the software

and regardless of the triggered effects or the presumable intent of an attacker. This also demonstrated that the technical aspects of cyberspace did not necessarily hinder arms control measures, as argued by Burgers and Robinson (2018) and Perkovich and Levite (2017), confirming the perspective of Reuter et al. (2022), that developing these tools for cyberspace is possible and demonstrating how it can be done. The analysis of the thesis also found that regardless of challenges that are still unsolved, like for example the attribution problem, further technological advances like the application of AI into cyberspace tools will likely increase the complexity for arms control debates. To a certain degree, the thesis found that these approaches can build on existing debates that have already been going on for some years for other types of weapons, for example for lethal autonomous weapon systems (LAWS).

Regarding the question of how existing IT measures can be used for cyber arms control and which adjustments or new approaches are necessary, the thesis shows that computer science already has developed a lot of tools and concepts, that can be adapted and applied to this context. In addition, the analysis highlighted that for many cases, the tools and infrastructures which are already used for IT security measures often provide the necessary tools that can be applied to arms control tasks, which falls in line with the initially mentioned premise of Altmann (2019b), to take advantage of synergies with existing IT measures, developed for civilian purposes. Based on this results, the thesis answers the question of how de-escalation and arms control measures for cyberspace can be developed by identifying suitable technical parameters and starting points. It develops and proposes such measures, demonstrate their applicability and discusses necessary adjustments of existing IT infrastructures as well as its limitations. Beyond that, the thesis provides impulses to connect these developed inter-state disarmament measures with already existing national approaches, like the consideration of vulnerability disclosure as argued by Schulze (2019) and US White House (2017).

Nevertheless, the thesis also discusses the political dimension of arms control and the constraints that come with this context. This especially relates to the fact that states presumably follow different, sometimes diverging interests, that will deviate from agreed upon behavior or circumvent measures if it's in their interest, thus requiring a great deal of pragmatism regarding the possible outcome of these measures. The thesis argued that this aspect can partly contrast the methodical approaches of a computer scientist, which usually try to find the optimal solution for a problem and where, from a security standpoint, a system is either secure or it's not. Nevertheless, given the presented challenges and necessity for tools that can help to foster the peaceful development of cyberspace, the thesis concludes that it is better to develop solutions that provide a small and limited gain of security, regulation or transparency for cyberspace than having no solutions at all. In light of the main research question of this thesis, these results match the overarching goal of arms control, to reduce the potential of violent conflicts while accepting the flaw and gaps in agreements. This perspective builds on the optimism that early small steps might make it easier for states to reach an agreement and comply with – as the necessary concessions are rather small – and that the progress made in this way improves the likelihood of future, more far-reaching agreements. The thesis therefore showed that, beside all limitations and necessary future work, it is possible to develop measures for the de-escalation of state-led conflicts in cyberspace and arms control of cyberweapons.

5.2 Limitations

The overarching approach to develop technical measures of arms control and conflict deescalation in cyberspace has some general limitations, which will be discussed in the following.

For one, the amount of related work that this thesis could be built upon is quite limited. Whereas the political dimensions of cyber arms control already have been debated quite extensively in political science as well as peace and security policy publications (e.g. Burgers and Robinson, 2018; Eilstrup-Sangiovanni, 2018; Futter, 2020; Hansel et al., 2018 or Altmann, 2019a), the actual application of existing IT measures and methods to provide technical solutions for cyber arms control is quite new and focus on the adoption of IT procedures like for example the usage of bug bounty programs as a measure of arms control (J. A. Silomon, 2020). Thus, the idea and results of this thesis of developing algorithms and software tools for arms control and de-escalation, as especially presented in chapters 16, 15 and 17 has to be seen as pioneering work. Although the constraints and requirements of the developed measures directly build upon political necessities and circumstances and the results are weighted against these, they still are part of scientific work that has yet to hold up in a real-world implementation. The thesis provides simulations and estimations where necessary in order to underline the applicability of the proposed measures (for example in chapter 16.7), but could at best provide proof of concept implementations (like e.g. in chapters 16.4, 16.5 and 16.6). Other parts, especially when it touches complex IT network infrastructures, still have to be implemented, modeled and evaluated (like e.g. discussed in 15.6.3). When it comes to treaty negotiations, especially the practical verification measures are usually discussed and agreed upon to an extremely detailed degree that defines the specific circumstance of the measure, the allowed tools, the access, the personnel to run or control the tools as well its evaluation and so on [1]. All of these aspects are part of the negotiation process and could be deliberately weakened, changed or modified in order to reach a level between all negotiating state that allows the adoption of a political agreement. It is, therefore, not necessarily useful to define the technical measure in every detail, but rather to demonstrate the effectiveness and applicability of the overall concept and highlight starting points (like for example in chapters 10.3 and 17.4).

Another limitation that inherently results from the political context of this work is the discrepancy between any measure – as waterproof and secure as it technically might be – and the actual compliance of the state to its hypothetical implementation. As discussed for instance in chapter 15.6.1, this aspect has always had to be taken into account by arms control measures. In the best case, the measure anticipates this discrepancy and conceptually mitigates these limitations (as for example discussed in chapter 16.3.1), non-complying behavior can be detected and politically sanctioned or the compliance itself offers such important benefits, for example through security guarantees or de-escalation options, that states refrain from deviating behavior in order to not jeopardize the effectiveness of the benefits as argued for example in chapter 15.6.3.

[1] As an example see the verification and monitoring regulations that are part of the Joint Comprehensive Plan of Action (JCPOA) (IAEA, 2015b), also known as "Iran deal".

A limitation that especially affects the performed expert interviews in chapter 9.4 lies in the fact that the development of technical cyber arms control measures requires a lot of highly specific technical knowledge of how cyberspace works, about IT networking concepts and techniques or about malware, their development, deployment and life cycle. Whereas cyberwar and cyberpeace have so far been the focus of political science and security policy debates, the sometimes insufficient detailed technical knowledge within these expert circles might have led to inaccurate or even false assessments of the feasibility of cyber arms control. This has especially become visible regarding the still ongoing perspective that attribution of cyberattacks is not possible (like discussed in chapter 11.1) or that cyberweapons cannot be defined due to the specific technical circumstances (as discussed in chapter 10.2); topics that have both been addressed with this thesis.

One final but important limitation regards the aspect that the debates on cyberwar and cyberweapons are very much subject to current developments and some concept that so far has only been discussed theoretically has to hold against actual events, for example the ongoing Russian war against Ukraine and the question of what constitutes a cyberwar as discussed in chapters 7.1 and 7.2. It is important to keep in mind that the sources to assess the events in this war or similar incidents are usually limited to news and other media outlets, social media channels or reports from IT security companies. All these sources are usually hard to verify, often lack technical details, and are sometimes meant to support the specific perspective of one side. This limits the scientific and empirical value of these sources, leaving room for interpretation.

5.3 Future Work

With regard to the results and the discussed limitations, future work should cover the refinement of the proposed measures and their modeling toward the preparation of real-world applications. This should include, for example, the simulation of IT networks with different topologies, sizes, activities and interconnections to create reliable estimations of logging sizes and required storage capacities regarding the proposed verification as well as the de-escalation measures. Although probably not implementable any time soon, this will further help to create a complete implementation blueprint as support for political negotiations. Regarding the global cyberspace infrastructures, the already existing model of the submarine cable network should be extended by including terrestrial cable connections, peering stations, and in the best case, satellite network supply to further enhance the analysis of dependencies, threats and weak points and bottlenecks of internet connectivity. Further work should also be put into the proposed and followed approach of applying existing IT methods and techniques to arms control, disarmament and de-escalation measures for cyberspace. This affects, for example, the question of marking civilian or humanitarian objects with a "digital red cross" and other symbols in order to highlight their special claims for protection, as agreed upon in humanitarian law for times of military conflicts. This work has already begun under the head of the ICRC (ICRC, 2022a). Although the mentioned limitations of non-compliance also apply here, this still could provide an aid to protect non-military IT systems from unintentional damages. Another possible future work regards the question of how the proliferation of cyberweapons can be tracked or regulated. Although such digital

products are seamlessly duplicatable, a lot of efforts have already been put into the comparable area of digital rights management (DRM) that tackles the same problem of restricting the usage of digital goods to its permitted – usually commercial – purpose. Finally, it will be important to discuss the above measures, ideas and concepts with political actors to test the applicability of the measures, refine them or develop new concepts that hopefully can be of support one day for political negotiations and treaties.

5.4 Conclusion

The thesis shows that computer science can provide input for the analysis of the specific challenges of arms control and de-escalation in cyberspace and that it is possible to develop technical solutions for this challenge. Although by far complete and merely an impulse for further research, the thesis presents and discusses food-for-thought impulses that can be built upon in future work and which counter the still widespread opinion that arms control in cyberspace is almost impossible. The thesis also shows and emphasizes the role and responsibilities of computer scientists and the necessity of their contributions to peace and conflict research in the natural sciences. Their know-how is of utmost importance in order to further develop tools and measures that will help to sustain the peaceful development of the domain cyberspace. Finally, a last but quite important conclusion that derives from the political aspects of this thesis is that it's essential and necessary to listen to the politicians and decision-makers that shape the political cyber realm. Although from a technical and engineering point of view often imperfect, this is the level where cyber arms control measures have to be presented, where they will be discussed and – in the best case – proceeded to international political forums that can pave the way towards a peaceful development of cyberspace.

PUBLICATIONS
PART A

6

FROM CYBERWAR TO CYBERPEACE

ABSTRACT The encompassing trend of digitalization and widespread dependencies on IT systems triggers adjustments even in the military forces. Besides necessary enhancements of IT security and defensive measures for cyberspace, a growing number of states are establishing offensive military capabilities for this domain. Looking back on historical developments and transformations due to advancements in military technologies, the chapter discusses the political progress that has been made and tools that have been developed since. Both of these have contributed to handling challenges and confining threats to international security. With this background, the text assesses a possible application of these efforts to developments concerning cyberspace, as well as obstacles that need to be tackled for it to be successful. The chapter points out political advancements already in progress, the role of social initiatives, such as the cyberpeace campaign of the *Forum InformatikerInnen für Frieden und gesellschaftliche Verantwortung (forum of computer scientists for peace and societal responsibility)*, and potential consequences of the rising probability of a cyberwar regarding the prospects of cyberpeace.

ORIGINAL PUBLICATION *Reinhold, T., & Reuter, C. (2019b, March 13). From Cyber War to Cyber Peace. In C. Reuter (Ed.),* Information Technology for Peace and Security: IT Applications and Infrastructures in Conflicts, Crises, War, and Peace *(pp. 139–164). Springer Vieweg. https://doi.org/10.1007/978-3-658-25652-4_7*

6.1 INTRODUCTION

In Iran in June 2010, a malicious software (**malware**) had been discovered on specialized industry control computers of a uranium enrichment plant, which has been used to sabotage the facility via centrifuge manipulation. Analyses of the program, which is now known as Stuxnet revealed that the sabotage had already been running for several years, and that the hackers must have possessed remarkable technical skills as well as detailed knowledge of the plant's construction. Because of the high development costs and effort for such malware capable to attack an industrial facility disconnected from the internet, a governmental agent was assumed to be the driving force behind Stuxnet. This assumption has been confirmed, and Stuxnet is now known to be a joint project of US and Israeli military and intelligence services (Nakashima & Warrick, 2012; Sanger, 2014). However, Stuxnet has not been the first malware allegedly applied by a state. For example, in 2007, the Israeli military was accused of sabotaging Syrian air defense systems (Fulghum, 2007). And in Estonia, servers have been attacked and temporarily disabled, presumably by Kremlin-based activists from Russia (Bright, 2007) – incidents which are said to have occurred during the Caucasian war in 2008 in a similar

form (Danchev, 2008). Since 2010, such events have been repeatedly receiving public attention, the latest case being in 2015, when the German Federal Parliament's internal communication system "Parlakom" was spied upon for months, and documents, access details and personal communication by deputies and their employees were presumably stolen. The attack severely impeded the parliament's work and could not be stopped until the system was shut down completely during the summer break. A video made by FIFf (FIfF e.V., 2017) motivates the discussion around **cyberwar** and **cyberpeace**. Their central argument why cyberwar needs to be prevented, and offensive cyber strategies of militaries and secret services stopped, is that cyberweapons are in many ways as dangerous and inhumane as biological and chemical weapons, which have already been outlawed by the international community. Accordingly **cyberweapons** are malware (such as viruses, worms and Trojans), which work only when based on loopholes in the security of alien systems. Therefore, cyber armament consist mainly of searching alien networks, institutions and devices for potential vulnerabilities, or even creating them. Of course, as there is a market for everything, access to and knowledge of security gaps can also be bought. In cyberwar, aggressors use their control over systems to harm the opposing party. In practice, this means that anything containing a computer can be attacked. Thus, every PC, every router and telephone, every control system, be it small or large, become potential weapons. If our **critical infrastructure** (e.g., transportation systems, waterworks, hospitals and power plants) were switched off or even used against us, the consequences would be just as devastating as in an attack with conventional weapons.

Nonetheless, governments around the globe are arming for offensive cyberwar and even Germany started to establish dedicated military cyber forces. A broad societal discussion about the legality of turning our devices into weapons that can be used against us at any time, has yet to materialize. However, FiFf names several reasons why cyberweapons should be outlawed, and money currently spent on keeping critical infrastructure vulnerable used to close security gaps instead.

- Firstly, **cyberweapons can be used anonymously**. In global virtual networks such as the internet, it is impossible to identify the real perpetrator, as they mostly use several devices to execute the attack in order to make backtracking impossible. Furthermore, they are often committed at a time that suggests a different origin. And even if traces of the attack can be found, they do not prove anything because they are digital, and it is therefore impossible to tell whether they were left intentionally or accidentally.

- Secondly, **cyberweapons cannot be controlled**. Malware is often programmed to have an independent existence. If it is then willingly used as a weapon, or simply activated by accident, is out of control. Weapons of this sort can lie dormant in systems for years before causing any harm. What further distinguishes cyberweapons from conventional weapons is that they can easily be stolen, infinitely reproduced and spread simply by copy and pasting them.

- Lastly, **cyberweapons are expensive and threaten us more than they benefit us**. Militaries and secret services spend vast amounts of money on analyzing systems and buying security gaps. As only open loopholes can be used as weapons, buyers of information on them have an interest to keep them open as long as possible. Consequently, huge quantities of money are being spent globally to deliberately

> keep our critical infrastructure insecure and vulnerable. Naturally, these weaknesses also can be, and are daily being, found and exploited by criminals and terrorists (FIfF e.V., 2017).

This chapter first illustrates the relevance of cyberwar as a realistic part of future warfare and goes on to identify current challenges that the militarization of cyberspace poses. A central difficulty consists of the application of international law to cyberspace, which is partly due to the characteristics of cyberspace and partly due to the lack of international norms and definitions concerning cyberspace. These problems also make arms control in cyberspace more difficult than of more conventional weapon types. We further present measures that could be taken towards achieving cyberpeace, and some campaigns that try to raise public awareness of the necessity to act in this direction.

6.2 Current Challenges of Cyberwar

6.2.1 *Militarization of the Cyberspace*

Since the discovery of Stuxnet, the term *cyberwar* – derived from the term *cyberspace* – has been coined in connection to incidents of this kind. However, it is neglecting an important distinction which has to be considered when handling and interpreting such events: If the initiators of a cyberattack have not been ordered directly by a government, the attack in question is a "normal" criminal offense, which is a matter of national and international criminal prosecution and police cooperation. For these, multilateral agreements already exist, such as the Budapest Convention on Cybercrime issued in 2001 (EU-Council, 2001). Only once a government is the assumed attacker, interpretation of the incident concerns the political level and becomes relevant in terms of international law.

Here, a critical distinction has to be made regarding an appropriate reaction: Are we dealing with an intelligence service **espionage**, **sabotage**, or military activities directed towards clear strategic goals? For this purpose, we need to look at the damage already inflicted. Depending on the attacker's intention and applied malware, the range can reach from simple theft to temporary shutdown of an IT service to a specific damaging of IT and subordinated systems (G. D. Brown & Tullos, 2012).

Questions concerning cyberwar are exceeding the purely technical aspect of IT system maintenance or attacks on such systems. Apart from the aspects of defense and offense, as well as the necessary tools, states' security-political and military-strategical doctrines play a significant role. These determine to which degree a state identifies the cyberspace as a military domain, and how it treats according measures by other states.

For a few years, since the discovery of Stuxnet at the latest, governments have been increasingly perceiving the cyberspace as a military domain. According to a study by the United Nations Institute for Disarmament Research, at least 47 states operated military cyber programs in 2013, of which ten nations had a nominally offensive intention (UNIDIR, 2013). Documents from Edward Snowden's collection give further evidence. We find that in 2012 Barack Obama, being US president at the time, instructed

his military and Secret Service leaders to create a list of the most important potential military targets in cyberspace and to develop solutions for the disturbance of these targets up to their destruction (Guardian, 2013). The consequence of this presidential directive becomes evident regarding the cyber espionage and manipulation opportunities revealed in 2013, which the National Security Agency (NSA) has been developing in the US, and partially distributed as hidden digital sleeper agent in commercial projects. Traditionally, the NSA is subordinated to the US cyber command leader, i.e. the offensive cyber forces of the US army, who therefore have direct access to NSA technologies. Since 2016, these have been officially used for the first time in the war against the "Islamic State" (US White House, 2016). In the Warsaw Summit Communiqué in 2016, the NATO has integrated defense in cyberspace into collective defense according to article 5, and is therefore also evaluating cyberattacks and the aspect of military aggression.

Germany has adapted to the change as well; the Federal Armed Forces already had a unit for Computer Network Operations (CNO) since 2006, that consisted of approximately 60 members. The CNO forces are assigned to the organizational unit of the strategic reconnaissance command. This unit's task is the offensive access to foreign IT systems. However, they are currently training in enclosed training networks, and have not yet been utilized according to official announcements (German Federal Parliament Defense Commitee, 2016). At the end of 2017 the Federal Defense Ministry has officially comprised the organizational units in the Federal Armed Forces that are dealing with IT and cyberspace into a separate organizational unit. "Cyber and information space" consists of 13800 offices and shares an organizational level with the military service branches of army, marine, air force as well as the medical service (German Ministry of Defense, 2016). Furthermore, the CNO unit has been enhanced to a "Center for network operations" and expanded by 20 posts. Due to the necessary intelligence information on relevant targets in the cyberspace, it is presumably cooperating closer with the federal intelligence service. The strategic guidelines of the White Paper show that these restructuring measures are linked to improved defense possibilities, as well as an enforced strategically offensive orientation of the Federal Armed Forces in cyberspace: "The empowerment of the Federal Armed Forces' common action in all dimensions is the superior benchmark" and an "impact superiority has to be reached across all intensity levels" (German Government, 2016) (translations by author). To reach this goal, the Federal Ministry of Defense in cooperation with Federal Ministry of the Interior, Building and Community founded a new agency for innovations in IT security that should take an example of the US Defense Advanced Research Projects Agency (DARPA). The tasks of this agency are to initiate, promote and finance research and innovation projects in the field of cybersecurity, especially "tomorrow's IT security solutions" (Manthey & Fleischer, 2023). For the period from 2019 to 2022, the agency can spend a total of around 200 million euros.

The increasing militarization of cyberspace holds a number of challenges in the domains of international law and security policy for the international society and individual states, which will be referred to in the following sections.

Until now, there has been no full-blown cyberwar. However, as mentioned above and in further detail in Table 6.1 below, there have been quite a few **cyber incidents** with different objectives and magnitudes. This hints at possible scopes and consequences of future **cyber attacks**, and therefore the (growing) relevance of the topic.

Year	Alleged actor[1]	Description
2007	Russia	The cyberattack on websites of the government and other institutions, banks and ministries of Estonia that prevented their access is often considered to be the first significant state driven cyberattack. An official involvement was denied by Russia and the attack attributed to a patriotic russian youth organisation.
2008	Russia	The cyberattacks performed against websites of Georgia and South Ossetia during the military conflict with Russia prevented public information platforms and media services from working. These incidents are often considered to be the first attempts to use cyber capabilities as measure in military conflicts.
2010	USA / Israel	The malware Stuxnet was used to silently sabotage the Iranian nuclear program. Its presumably long development and deployment time, which involved very specific information on the targeted industrial systems, were an international "eye opener" how states use attacks over the cyberspace for foreign policy intends.
2012	Iran	A malware named Shamoon/Wiper was used against industrial oil companies in Saudi Arabia. The malware had been explicitly developed to spread out fastly within infected networks and to render the targeted computers useless by deleting relevant operating system files. It affected up to 30.000 IT systems.
2012	USA / Israel	The malware Flame was used for espionage and intelligence purposes in the Middle East. It was considered to be the most versatile malware development so far with a huge variety of modules to infect different IT systems and perform multiple tasks on them. Therefore, Flame is seen as the first state developed "cyberattack multi purpose frame work".
2013	China	A report from the US-based IT security company Mandiant analysed several long term cyberattacks and revealed a military cyber force in China, based on IT forensic analysis. The Unit "PLA 61389" had been accused of different espionage attacks with custom tailored cyberweapons.
2014	Israel	The malware campaign Duqu 2.0 was used for espionage purposes with particularly versatile cloaking mechanisms. It is presumably a further development and extension of earlier versions that had been detected 2011.

[1]The alleged actor is mostly based on published information by intelligence or law enforcement agencies. The underlying evidence had been seldomly revealed and it had to be taken into account, that such charges can have political motivation too. Also it's important to note, that the distinction between hacking activities by a state and its institutions and non-state groups that are not directly connected to a state but under its indirect control is hard to make.

2014	Palästina	XtremeRAT was a spear phishing malware campaign in the context of the Middle East conflicts that had been used by a Palestinian activist group for espionage and data theft.
2015	USA	The Equation Group is the name of a malware campaign with an extremely complex infrastructure and technological basis. The campaign had been active for several years, with earliest indications from 1996. Its highly developed tools and malware frameworks had clearly been developed and extended over years and share similarities with incidents like Stuxnet and Flame.
2015	Russia	In the context of the western Ukraine conflict, Russia was accused for attacks against Ukrainian energy companies that stopped the power supply for around 700.000 residents for several hours. The malware Black-Energy and Killdisk were used to gain access and shut down IT systems.
2016	Russia	In the preparations of the US presidency elections of 2016, cyberattacks were performed against the Democratic National Committee that led to a severe data breach. Some of the documents were leaked subsequently. The cyberattack is seen as part of severe and long-lasting interference within the democratic election process of the USA.
2017	Iran	A malware that targeted specific industrial control systems (SCADA) was deployed against Saudi Arabian petro chemical companies. It had been specifically designed to trigger physical harm and destruction in these facilities, although this never happened due to programming errors.
2017	North-Korea	After the leak of the fatal zero day exploit EternalBlue, wich had been stolen from the NSA and affected Microsoft Windows systems, a malware called "WannaCry" was deployed that used this exploit. It could have spread massively around the world and hold affected users to ransom by encrypting their hard drives.
2018	Russia	In spring 2018, a hacking attack against German governmental IT systems and networks was published. The attack had been active but cloaked for more than a year and had been performed very carefully - without automatic replication or infection of IT systems. Its primary goal presumably had been espionage.

Table 6.1: List of relevant cyber incidents with presumably state or state influenced actors
Source for all: https://cyber-peace.org/cyberpeace-cyberwar/relevante-cybervorfalle/

6.2.2 *International Law in the Cyberspace*

With regard to the established rules of international operation, the question arises how they can be applied to cyberspace. The difficulty of this debate already becomes evident

with the discussions on a common definition of cyberspace: While the US-American and Western European interpretation is guided by technical standards and covers the number of IT systems and their network infrastructure so that security mostly refers to the integrity of these systems, other countries like Russia or China consider the information which is saved, transmitted and published therein as part of the cyberspace. As a result, security, especially on a national level, exceeds the integrity of technical systems and becomes an issue of control of and access to this information – a point of view which is difficult to reconcile with human rights principles (UN General Assembly, 2011).

TALLINN MANUAL The experts of the NATO Cooperative Cyber Defense Center of Excellence (CCDCOE) made a first effort towards solving this problem in 2013 with the so-called "Tallinn Manual", a handbook including 95 guidelines for nations in case of a cyberwar. Even though it is not binding, it points out the specific characteristics of the cyberspace in which international law applies (CCDCOE, 2013), and indicates how international law can be interpreted for military conflicts in this new domain. In 2017, the CCDCOE published a second version of the manual called the "Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations" (CCDCOE, 2017) that continues this evaluation, especially of state behavior, as well as rules and norms in peace time.

VIRTUALITY OF CYBERSPACE The central challenge lies within the virtuality of cyberspace, which undermines approaches and regulations based on territorial borders or the localization of military means. Equally problematic are the immateriality of malware as well as the unlimited possibility to reproduce it. Furthermore, due to the structure of cyberspace and the principles of data transmission, it is easy to act secretly or to cover up the actual origin of the attack. In addition, IT systems are often highly interconnected, and directly or indirectly control processes of so-called critical infrastructures, such as electricity or water supply, communication or traffic (German Ministry for Interior Affairs, 2009). The impairment of a nation's IT system can therefore have potentially incalculable consequences of grave impacts on originally not intended targets. Because concealed access to IT systems with the aim of espionage or military situation overview is often linked to the application of malware and manipulation of the IT system functions, the threshold for such threats is very low.

Regarding central concepts of international law, these characteristics of cyberspace raise a range of issues. For example, this concerns the international agreement on nonviolence and the right of self-defense according to article 2, paragraph 4, and article 51 of the UN Charter, as well as the **principles of adequacy** and **proportionality** of military reactions: What does "use of force" mean in the cyberspace? When are malware and diverse cyber attack tools and methods considered "weapons"? When do we speak of an "armed attack"?

Previous approaches to apply these concepts to the cyberspace usually refer to consequences of classical, so-called kinetic weapons to evaluate specific cyber incidents and possible reactions legitimized by international law. Thus, the Tallinn Manual defines armed attacks in cyberspace as "cyber activities that proximately result in death, injury, or significant destruction" (CCDCOE, 2013).

CHARACTERISTICS OF THE APPLICATION OF MALWARE Such an approach, however, falls short since it does not sufficiently consider that the scope, timing and form of damage of cyberattacks are in many forms not comparable to conventional weapons:

- First, it is possible for malware to spread uncontrollably beyond IT networks and affect external systems which were not the target of the attack and which possibly belong to an uninvolved nation. For example, inactive versions of Stuxnet have been discovered on tens of thousands of systems worldwide (Falliere et al., 2011). Application of malware operating secretly over a longer time frame or using indirect ways of sub-system manipulation, and thus not inflicting directly visible and assignable damage, is equally problematic.
- In addition, the current trend towards cloud technologies further complicates the geographical localization of IT systems. Linked to this is the so-called **attribution problem**: Every nation's right of self-defense implies that the origin of an attack to which the nation is forced to react promptly, must be clear. In cyberspace, however, as mentioned above, it is common practice to carry out attacks from external systems specifically hijacked for this purpose to cover up the source. As a consequence, the retracing of these attacks through several steps cannot be carried out timely and in a forensically secure manner. The precise limitation of permitted military use of malware proves to be equally difficult. Usually, IT tools, methods and software used by criminals, IT security experts and military forces to access IT systems are barely distinguishable. Nevertheless, depending on the intention, their usage has very different outcomes: E.g., revelation, analysis and remedy of weaknesses (IT security expert), theft of credit card details (criminals) or the destruction of an air force monitoring program (military). Apart from the tools, the identifiability of state or military agents and the term combatants in cyberspace, as well as their distinction to civilians, are hard to achieve with current technologies. However, such labels are essential for dealing with agents in crisis and war situations.

In the United Nations and the Organization for Security and Co-operation in Europe (OSCE), expert groups are discussing these questions. However, we cannot yet see specific approaches for binding international regulations in cyberspace, especially with regard to the "right to war" (***ius ad bellum***) and the "law of war" (***ius in bello***).

6.2.3 *Lacking International Norms and Definitions*

CYBERWAR VS. CYBERCRIME A basic problem when evaluating incidents in the cyberspace consists in the distinction between normal criminality in cyberspace, so-called **cybercrime**, and governmental actions as well as those directed against other nations, referred to as cyberwar. Furthermore, the evaluation of a threat by a cyber incident as well as the reaction on political and legal level is up to the affected state. Based on already established regulations on cybercrime, international agencies like ICPO-Interpol or Europol are dealing with international criminality in cyberspace. At the same time, the European Network and Information Security Agency (ENISA) is consulting and connecting EU states via cooperation centers.

In contrast to this, it is difficult to apply established norms to cyber incidents which are allegedly traced back to state agents or third parties under governmental order, since the partaking agents cannot be identified and therefore covenants cannot be verified, and because of a lack of internationally binding agreements. It is controversial whether international humanitarian law can be applied to the cyberspace because of national sovereignty and the right of self-defense, but also with regard to nations' responsibilities in cyberspace. Another question concerns the scope of damage caused by a cyberattack, which would correspond to an armed attack and therefore legitimize national self-defense according to art. 51 of the UN Charter.

The NATO CCDCOE, among others, has been largely contributing to the answer to these questions with the two Tallinn Manual publications (CCDCOE, 2013, 2017), along with the UN Group of Governmental Experts with their reports (Tikk-Ringas, 2012) and the Organization for Economic Co-operation and Development (OECD). All are dealing with the application and extension of established norms of international law to the cyberspace, difficulties and limitations resulting from this, and discussing different solution approaches. While the groups agree on the fact that cyberattacks, under certain circumstances, can violate the national sovereignty, there are significant differences concerning clear definitions for cyberattacks. Especially so, when it comes to their comparability to armed attacks and the issue of appropriate reaction to a cyberattack, such as the use of conventional weapons. The underlying differences of states on these issues still strongly inhibit the development of internationally binding agreements (Tikk & Kerttunen, 2017).

Binding Norms Apart from questions concerning the motivation for a cyberattack, establishing binding norms is further complicated by differentiating between cyber activities without the intention of damage, and those attacks which are actively carried out with the aim to disrupt external IT systems. Both kinds of access basically correspond to similar principles and use comparable tools. They particularly differ in terms of the malware installed and controlled by the attacker, which performs the desired damaging function on the target system (**payload**). The latter can consist of copying and stealing information, but also in completely shutting down thousands of afflicted PCs, as demonstrated in the attack on the Saudi company Aramco (Bronk & Tikk-Ringas, 2013).

Attribution Problem Another problem for applying international law lies within the attribution problem of attacks in cyberspace mentioned above, i.e. timely identification of an attack source. This is much harder in the cyberspace than with conventional weapons, since the attackers possess a great range of options to cover up their own identity. Even though debates often refer to the practical impossibility of attribution, authors like Herb Lin (Lin, 2011) argue that under certain circumstances, the identification of the origin network is sufficient to gain details about the offender, so that the exact source computer does not necessarily have to be identified. Apart from this, the planning and operation of a specific access to complex systems takes a certain time, where transmission data can be collected, forensically analyzed and used for an attribution under consideration of the current international political situation (Clark & Landau, 2010). Using this approach, in spring 2013, the US-American IT forensic company Mandiant identified a cyber unit of the Chinese People's Liberation Army

(PLA Unit 61398) as initiators of several attacks against US-American organizations and institutions carried out throughout many years. They published their insights (Mandiant Corporation, 2013) at a time of high-level meetings between the US and Chinese presidents and state secretaries on security in cyberspace.

Elaboration of International Norms and Cyber Weapons Furthermore, the elaboration of international norms for the cyberspace becomes difficult due to the aforementioned definition of cyberweapons. As explained above, the hardware and software tools for accessing external systems do not reveal many details on the specific intention. The OECD analyzed this question with regard to characteristics of conventional weapons: *"There is an important distinction between something that causes unpleasant or even deadly effects and a weapon. A weapon is "directed force" – its release can be controlled, there is a reasonable forecast of the effects it will have, and it will not damage the user, his friends or innocent third parties."* (Sommer & Brown, 2011). Based on these criteria, the authors of this OECD study identified important reference points for the evaluation of specific malware, taking into account technical details, the political situation of the national agents, and their presumed intention. They suggest a classification of all malware in a continuum between "low level cyberweapons" (the manipulation of websites or purposefully sent emails inflicted with malware for espionage purposes) and "high level cyberweapons" (attacks with direct and lasting disturbing or destructive effect). A sufficient distinction of malware and the decision whether it is a weapon according to international law can therefore only be made in the context of individual cases.

6.2.4 *Arms Control in the Cyberspace*

The presented difficulties and ambiguities which the international community is facing with regard to militarization of the cyberspace also raise issues of security policy. On the one hand, considering the increasing cyber threats and the higher awareness of risk around critical infrastructures, it is clearly important to protect IT systems more effectively and sustainably. On the other hand, improvement of defense know-how, analysis of attack scenarios and identification of weak points also imply an increase in potential ability for offensive actions in IT systems. A sensible technical distinction is not possible at this point, while limitations to purely defensive activities by military forces are of declarative character only.

Active Defense Similar problems emerge from the **active defense** concept considered by NATO CCDCOE (Minárik, T. & Stinissen, Jan, 2014) and German Federal Armed Forces (German Federal Parliament Defense Commitee, 2016). The essence of this idea lies within preventing cyber threats not only by purely defensive measures like disconnecting network connections, but also via **hack-back**, i.e. the intrusion into and disruption of the offender's IT systems. Apart from the problem that the perceived source of an attack does not necessarily lead back to the actual attacker, offensive capabilities have to be established here. Furthermore, an elaborate knowledge of the domain is required, i.e. knowledge of the goals, their state and technical details, as well as on the used software and its version, to be able to use cyberweapons effectively and

purposefully, so that, if necessary, intelligence service activities can be initiated in the potential attackers' IT systems prior to an attack.

Apart from this, knowledge on security gaps in the target systems are necessary for specific access. In many past incidents, security gaps in popular and widely used software such as email programs, browsers or Office applications have been used. An increase in military offensive activities does not benefit an open approach to security gaps and their closure – instead, the trade with such knowledge has been flourishing, be it on the black market or by companies that seek, buy and commercially exploit such security gaps (Reinhold, 2014a).

Dual Use Along with the militarization of cyberspace, considering the current uncertainties on the international evaluation of the new military potential, there is a risk of an arms race between states. With regard to the established international arms control measures and disarmament initiatives, new questions arise in this context. IT assets as well as software security gaps with potential military value are commonly used by civilians. While this so-called **dual use** character creates the necessity for a thorough export examination, the software characteristics mentioned above make it difficult to comprehend the proliferation and use, cases of exports and to verify the commitments of importers and purchasers of these systems.

As a first step for monitoring trade with IT systems of value for intelligence service or military, the Wassenaar Arrangement on Export Controls for Conventional Arms and Dual-Use Goods and Technologies, established in 1995, has been extended to include so-called intrusion software in 2013 ("The Wassenaar Arrangement on export controls for conventional arms and dual-use goods and technologies - List of dual-use goods and technologies and munitions list," 2017) (The Wassenaar Arrangement Secretariat, 2022). Even though this multilateral arrangement which currently includes 42 states should be regarded critically (Holtom & Bromley, 2010), it is an essential starting point for establishing regulations and the future of arms control in cyberspace.

In order to prevent an arms race, further trust-building measures between states are crucial. These should allow states to discuss their ideas of security, perceived threats and those addressed in the context of security strategies, as well as initiated measures. The goal is "to reduce and even eliminate the causes of mistrust, fear, misunderstanding and miscalculations with regard to relevant military activities and intentions of other states" (UN, 1988) and to establish communication channels for further conversations or crisis situations.

First bilateral agreements on a common interest in security of civil IT systems, as well as limitation of potentially threatening intelligence service espionage, already exist. Especially the USA and China have been leading high-level discussions in the past years, establishing the first bilateral contract specifically referring to IT security in 2015, where both states addressed important potential threats in the cyberspace (Nakashima & Mufson, 2015). This process has been accompanied by bi- and multilateral military crisis training for cyber incidents (Hopkins, 2012).

Computer Emergency Response Teams Another important step towards trust-building measures consists in the development and establishment of collective incident

reporting systems, i.e. clearly structured and hierarchical warning and reporting systems for critical cyber incidents, such as already existing **Computer Emergency Response Teams** (CERT) on national level, or for partial networks like academic research associations. The European Union is moving towards a transnational protection of IT infrastructure stability by introducing national obligation to report such incidents, and an interconnected transmission crossing national borders.

All this is contributing to reducing irrational fear of the "cyber doomsday" which is often spread through media. The cyber incidents of the past years have shown that cyberattacks by state agents rarely result in total conflicts carried out over the internet, but rather become of interest for foreign policy, as it is the case with classical espionage incidents. For example, the US government used a data theft in the context of a cyberattack on an affiliated company of Sony located in the US in 2013 as an opportunity to impose sanctions on North-Korean citizens and companies, even though there was no sufficient evidence.

6.3 Measures for Cyberpeace

The militarization of cyberspace also concerns its civil, individual use. The NSA affair of 2014 and 2015 has demonstrated the wide range of surveillance and control options in the cyberspace – from an aggregation of various data by IT services and social networks to total surveillance or a well-aimed hardware manipulation (Appelbaum et al., 2013) – and the degree to which their military use in the context of international competition for dominance in cyberspace affects universal human rights. The destructive and economically disastrous malware campaigns WannaCry and NotPetya from 2017 (Ehrenfeld, 2017; Fayi, 2018; Fruhlinger, 2017, 2022; Mohurle & Patil, 2017), both based on zero days exploits which had been stolen from the NSA, demonstrated once again the risks of the non-disclosure of vulnerabilities for intelligence or military purposes.

At the same time, the cyberspace resembles the commons regarding its broad impact and social dependencies (Ostrom, 1990). Constant intelligence service activities in the cyberspace as well as the purposeful weakening of IT systems, or the conscious manipulation of IT infrastructures in favor of military strategies are hence impairing a commonly used asset.

Therefore, it is essential that states face the numerous challenges on the way to a peaceful use of cyberspace. Apart from the aforementioned questions referring to arms control and trust-building measures, these challenges also concern the structures behind the cyberspace itself: The discussions about an increased participation by international committees such as the International Telecommunication Union of the United Nations in decisions concerning the development and technological expansion of the cyberspace are still ongoing. For quite some time, emerging nations like Brazil have been demanding an end of dominance of the US-American Internet Corporation for Assigned Names and Numbers, which is coordinating the domain name system and the assignment of IP addresses, as well as a broad participation of all nations in designing the cyberspace. However, even economic actors that often provide the technical infrastructures or

essential services demand multi-stakeholder debates on the future embodiment of the cyberspace and binding rules for the actors in this domain[2].

As a domain defined and controlled completely by humans, the cyberspace offers prerequisites for a peaceful formation on the one hand – that is, if we are successful in establishing an international sense for its importance. On the other hand, the all-destructive cyberwar will probably never happen due to increasing international dependencies. “Cyberweapons” will rather be included in the military strategic planning arsenal, and primarily used along with conventional methods. However, this should not satisfy all peace activists.

Due to different characteristics of problems cyberwar and cyberpeace pose, as well as the multitude of stakeholders involved and their interests, various possibilities to influence and shape the process are offered. To do this successfully, measures need to be targeted at the respective bargaining level and context of discussion. In this context, Götz Neuneck (Neuneck & Mölling, C., 2001) proposes differentiating between three areas of measures:

- **cooperative** and **declaratory approaches**
- **informational approaches** and
- **technical approaches**

  In the following these areas will be presented. As cyberspace provides the unique chance of perfect human control and design, the focus of information scientists lies on questions regarding the possible realization of peace building measures, such as **trust building**, **arms control** and **verification** by technical means. To be more precise, they consider how technical foundations and operating principles of cyberspace can contribute to this goal. Although findings from past decades concerning similar lines of questioning in different technological areas (e.g., nuclear armament, biological and chemical weapons, as well as the Outer Space Treaty) are not necessarily transferable, the experiences of these long-standing endeavors can provide important indications and impulses for the upcoming debates on the peaceful usage of cyberspace.

6.3.1 *Cooperative and Declaratory Approaches*

Cooperative approaches pursue coordination and trust building at a low level amongst relevant actors of the different states and their military organizations. In practice, this implies promoting interaction of representatives at conferences and in workshops. While doing so there is opportunity to discuss and explain threat scenarios, cyber doctrines and security concepts, in order to gain mutual and common understanding of the problems, as well as develop a uniform language regarding the issues at hand. Moreover, joint military drills to cyber scenarios can help to establish channels of communication,

[2] As an example, see the proposal for a "Digital Geneva Convention" by Microsoft (https://blogs.microsoft.com/uploads/2017/03/Transcript-of-Brad-Smiths-Keynote-Address-at-the-RSA-Conference-2017.pdf) or Google’s proposal for a new law framework (https://www.blog.google/topics/public-policy/digital-security-and-due-process-new-legal-framework-cloud-era/)

reduce worries of armament and mistrust. Examples of such drills are "Cyber Europe" (2010 and 2012) (ENISA, 2011, 2012) and the China-US-Wargames (2012) (Hopkins, 2012), the latter of which were organized by NGOs.

Another possible approach consists of establishing platforms for the purpose of exchanging information on the details of defensive and offensive measures the respective actors are conducting or planning in cyberspace. Such information can compensate perceptions of opposing parties' potential for aggression and destruction. Emergency communication could also be conducted over such channels, which can serve as an early warning system in the way of the 'red telephone'.

Further cooperative approaches are mutual support ("capacity building") in establishing national measures of protection against cyberattacks, linkage of national reporting- and emergency teams for cyber incidents (CERT), the development of collective cyberspace treaties, and in the long run, measures of arms control and verification. Particularly for the latter, however, there is an apparent lack of willingness to cooperate.

Next to these cooperative approaches, there are declaratory ones that states can unilaterally self-commit to as a **policy of détente**. Among these are the defensive orientation of armed forces as well as their security- and defense doctrines, and limitations in terms of the establishment of cyber forces. This can be reflected in the total amount of cyber forces, their drills and training scenarios, their technical equipment and organizational embedment in military operations. Renunciation of the "first use" of cyberweapons also belongs into this category.

A large fraction of these measures is of regulative character. It is in the nature of rules that they are, inter alia, declared out of political rationales and can be broken. Nonetheless, they are suited to counteract distrust, misjudgment of opposing parties' potentials and motivations, and rash reactions.

6.3.2 *Informatory Approaches*

A substantial part of states' security concepts comprises the collection, central notification and analysis of security incidents in state-owned and commercial institutions. In the realm of cyberspace, the concept of CERTs has existed for several decades. These central, intra-organizational registration offices collect incidents and report them to affiliated CERT-organizations, to warn and inform partners about security problems. This concept is being picked up by states for some years now, and extended, linked and hierarchically organized in whole economic branches up to government agencies. Especially the European Network and Information Security Agency (ENISA) promotes such linkage inside and between EU states and develops concepts for the categorization of cybersecurity incidents (ENISA, 2018), as well as the classification and definition of security warning levels (Dekker & Karsberg, 2014).

A further measure in this area is the creation and harmonization of statutory reporting obligations of relevant security incidents in the commercial and private sector, in order to identify cyber threats in good time and share this information over CERT infrastructures.

6.3.3 *Technical Approaches*

As mentioned above, the development of peace building technical options is an important part of this project. Such measures are currently barely being discussed on an international level, although the technology of cyberspace is firstly designable, and secondly, a multitude of relevant data and information that are suited for interchange and transparency building are already generated and saved by computer systems. The spectrum of technical measures that can be analyzed encompasses short-term approaches from the field of classical cybersecurity, such as the exchange and analysis of communication and log data of computer systems and networks, as well as more research-intensive questions, such as the improvement of the detectability of cyberattacks and their origin, or questions of mapping borders and the attached accountability of states in cyberspace (so-called attribution). Further aspects concern the depiction of neutral territory and objects as defined by the Geneva Convention, or the development of sensory measures of verification of cyberspace disarmament treaties (Reinhold, 2018a).

6.3.4 *Cyberpeace Campaign*

In their campaign "Cyberpeace" (FIfF e.V., 2017), the Forum calls to end all military operations on the internet by raising awareness of such dangers for, among others, individual privacy and human rights. The greatest danger, according to the Forum, lies in (unreported) weaknesses and loopholes inside IT systems which are used for cyberattacks. Because such attacks can hardly be controlled, they might affect civilian parties and even critical infrastructures responsible for the supply of energy, water, communication and health, and other IT systems with potential security gaps. Especially governmental cyberattacks, which possess most resources and influence, can weaken these systems and pose a threat to the functioning of society and even human lives.

The Forum demands that all cyberweapons be abolished by creating binding international arrangements on arms control, disarmament and the renunciation of developing and using cyberweapons for offensive actions on a governmental level. Meanwhile, the internet should function as a civil and peaceful resource without being misused for spying on civilians. Connected to this, the concept of general suspicion should be abandoned and replaced by achieving reliable evidence. The detailed demands can be found in Table 6.2.

The threshold for military activities is lower on cyber level as it does not create the impression of an actual war, which makes the abolishment of all cyberweapons necessary (see Table 6.2, demands 1, 2 and 3). This involves the extension of already existing agreements like the Geneva Convention to cyberspace (demand 5). Especially when it comes to critical infrastructures which guarantee the supply of existential goods and services, whose failure can threaten human lives, their disruption from outside should be treated as a crime of war (5). Any operator of critical infrastructures should be obliged to independently and transparently secure and protect their systems from attacks, and, if possible, detach them from the internet to prevent access for offenders (11). At the same time, governments should establish an internationally binding cyberspace initiative to

protect the internet as critical infrastructure and support the research and development of peace strategies (6).

The employment of conventional weapons as a reaction to a cyberattack equally runs counter to the Forum's peaceful policy. Because of the attribution problem, the source of a cyberattack cannot be clearly identified. Therefore, conventional weapons could cause a military escalation without a valid body of evidence (4).

Nonetheless, nations are urged to pursue a defensive strategy to protect their IT systems against cyberattacks, and therefore be allowed to use (hacker) tools for defense and exposure of existing security gaps (2 and 10). Such security gaps, once identified, should be officially reported, especially for public and corporate IT systems, and closed before they can be exploited, instead of leaving them open for intelligence services or armed forces (10). Consequently, public awareness of and trust in defensive cyber strategies will be raised. Furthermore, to prevent such weaknesses from emerging in the first place, security should be a central aspect for the architecture of computers, operating systems, infrastructures and networks (6, 11 and 14). Education around IT skills and their significance for society should be promoted to increase the number of qualified experts, improve security and quality of IT systems, and invigorate discussion on ethical and political issues around technology (13).

Transparency and democracy are further central aspects of the campaign. By officially promoting independent and transparent development, examination and risk analysis of software, loopholes can be openly identified and prevented, increasing security especially for critical infrastructures (14). Furthermore, instead of being the domain of secret services and military consultation companies, cybersecurity strategies and attacks should be officially confirmed and openly discussed with the goal to include them into the democratic decision process (7). As freedom of speech and assembly are basic human rights, they should be equally respected in cyberspace and not justify criminal prosecution or military activities (8). To further help protect human rights, independent and democratically regulated cybersecurity centers should be established that work towards preventing cyberattacks and establishing cyberpeace (12).

As an important tool for the formation of public opinion, discussion of the cyberspace in media and politics should follow defined terms and not be used to mislead and fuel conflict (9). Therefore, the Forum also offers definitions for a better understanding of cyberspace-related terms.

Demand	**Details**
1. No Pre-emptive or Offensive Strikes in Cyberspace	Nations should oblige themselves not to make offensive moves against others in cyberspace, while international agreements and cooperations on the prosecution of cybercrime should be extended to military and secret service activities.
2. Purely Defensive Security Policy	Instead of developing and using cyberweapons for offensive purposes, nations should apply a defensive strategy of protecting IT systems against cyberattacks.
3. Disarmament	Regulated by international agreements, nations should completely disarm on cyber level. This does not concern (hacker) tools for defense against cyberattacks and the exposure of existing security gaps.

Demand	Details
4. No Conventional Response to Cyberattacks	Because of the attribution problem, the source of a cyberattack cannot be clearly identified. Therefore, conventional weapons should not be used to respond to such an offense to prevent a military escalation without valid evidence.
5. Geneva Convention in Cyberspace	All applicable requirements of the Geneva Convention should be extended to cyberspace, and their disregard treated as a crime of war. This especially concerns critical infrastructures for the supply of existential goods and services, whose failure can threaten human lives.
6. Government-Level Cyberpeace Initiative	Governments should establish an internationally binding cyberspace initiative to protect the internet as critical infrastructure and support the research and development of peace strategies.
7. Democratic Internet Governance and Democratic Control over Cyber Security Strategies	Instead of being the domain of secret services and military consultation companies, cybersecurity strategies and attacks should be transparent, officially confirmed and openly discussed, with the goal to include them into the democratic decision process.
8. Online Protest is not a Crime	As freedom of speech and assembly are basic human rights, they should be respected in cyberspace and not justify criminal prosecution or military activities.
9. Clearly Defined and Demilitarised Political Language	Terms in the context of cyberspace should be officially defined and not used to mislead and fuel conflicts, as it currently is the practice in politics and media.
10. Obligatory Disclosure of Vulnerabilities	By officially reporting security gaps, especially for public and corporate IT systems, it should be ensured that these are closed before they can be exploited, instead of leaving them open for intelligence services or armed forces. Consequently, public awareness of and trust in defensive cyber strategies will be raised.
11. Protection of Critical Infrastructures	Any operator of critical infrastructures should be obliged to independently and transparently secure and protect their systems from attacks, and, if possible, detach them from the internet to prevent access for offenders.
12. Cybersecurity Centres	Independent and democratically regulated centers should be established to prevent cyberattacks, protect human rights and work towards cyberpeace.
13. Promotion of (rookie) IT Experts	Education around IT skills and their significance for society should be promoted to increase the number of qualified experts, improve security and quality of IT systems, and raise discussion on ethical and political issues around technology.
14. Promotion of FLOSS (Free and Libre Open Source Systems)	By officially promoting independent and transparent development, examination and risk analysis of software, loopholes can be openly identified and prevented, increasing security especially for critical infrastructures.

Table 6.2: Detailed demands of the Cyberpeace Campaign

6.4 Conclusions

The answer to the introductory question crucially depends on the underlying concepts of cyberwar and cyberpeace. These are open to discussion, as the disputes on definitions of crucial terms, such as cyberweapons or cyberspace, are unresolved. Consequently, in times of increasing militarization of cyberspace, applying international law to it is still challenging. At the same time, there are more and more activists who try to frame cyberpeace. Among them is the *forum of computer scientists for peace and social responsibility* who advocates international disarmament and purely defensive cyber military capabilities, as well as increasing formalization of organization and international law in cyberspace.

To recapitulate, the central challenges the cyber arms pose are:

- The militarization of cyberspace.
- Necessitated by its militarization, the application of international law in cyberspace. Difficulties result from the characteristics of cyberspace and malware (which lead to problems of attribution and therefore problems distinguishing cybercrime from cyberattacks), as well as the lack of international norms and definitions.
- Arms control in cyberspace, complicated by the above-mentioned problems. The offensive usefulness of defensive cyber capabilities and the dual use character of civil IT systems further impede efforts made.

Measures to overcome these problems and achieve cyberpeace include:

- Cooperative and declaratory approaches, i.e. promoting interaction and the exchange of information on the one side, and unilateral commitments to arms control on the other side;
- informational approaches, i.e. increasing cooperation when it comes to the collection of information; and
- technical approaches, i.e. increasing cybersecurity by technical means, especially by intensifying research.

Or, more programmatically put (by FiFf):

- Allowing purely defensive cyber policies only. The focus should lie on the protection of IT systems, all other capacities should be disarmed.
- Illegalizing conventional responses to cyberattacks. As the source of a cyberattack cannot be identified, conventional weapons should not be used in response.
- The extension of the Geneva Convention to cyberspace, in order to make state legally liable for their actions in cyberspace.

7

MILITARY CYBER ACTIVITIES IN RUSSIA'S WAR AGAINST UKRAINE AND THEIR SIGNIFICANCE FOR THE DEBATES ON THE CONTAINMENT OF A CYBERWAR

ABSTRACT Russia's invasion of Ukraine and the ensuing war have, among many other security certainties, demonstrated for the first time the role of cyberspace in an open war of aggression and revealed developments worth considering. The objective of this paper is to analyze military activities in cyberspace in the context of Russia's war against Ukraine based on publicly available information, and to assess them in terms of the previously prevailing notion of cyberwar as opposed to its actual role. Based on this, possible conclusions are considered, firstly regarding the future significance of cyber activities for Russia in times of peace and conflict in terms of the general military use of cyber assets, and secondly with regard to future international debates on the containment of cyberwar and the harmful use of cyber assets.

ORIGINAL PUBLICATION *Reinhold, T., & Reuter, C. (2023b). Zur Debatte über die Einhegung eines Cyberwars: Analyse militärischer Cyberaktivitäten im Krieg Russlands gegen die Ukraine.* Zeitschrift für Friedens- und Konfliktforschung. *https://doi.org/10.1007/s42597-023-00094-y* This text has originally been published in German. This is a translated version.

7.1 CYBERWAR - EXPECTATIONS

Even if terms such as cyberwar and cyberweapons remain controversial and no internationally binding definitions have yet been found (Reinhold & Reuter, 2021), the massive expansion of military cyber capacities worldwide in recent years and corresponding announcements[1] illustrate the importance that states and militaries attach to this domain. In the past, there have also been widely differing views on the role and extent of the specific use of military cyber assets in state-led military conflicts. With the Russian attack on Ukraine, there is now for the first time an example of warfare that also includes military action in cyberspace and which will be analyzed below against the background of previous expectations of this form of warfare.

The prevailing theories of anticipated state-led cyberwar to date can be placed on a continuum between a primarily intelligence-focused meaning of cyberweapons (Rovner, 2020) on the one hand, and the use of cyberweapons as a replacement for conventional warfare and boots on the ground (Bigelow, 2019) on the other. In the theories of the intelligence-focused importance of cyberweapons (Rovner, 2020) the domain of

[1] cf. Review analyses such as e.g. (Voo et al., 2020) or announcements by individual states, such as e.g. (Herpig et al., 2020) as well as (Lyu, 2019)

cyberspace serves primarily for intelligence and military information gathering, as well as tactical and strategic planning and command and control in warlike conflicts. Accordingly, such a form of cyberwarfare is characterized primarily by extensive but cautious and covert cyber operations that focus primarily on information gathering and - apart from unauthorized information leakage - do not pursue damaging or damaging intentions. Instead, in both peacetime and conflict, actors will seek to carefully protect source IT systems successfully infiltrated by cyberattacks and continuously cover traces to maintain access to stored information for as long as possible (Baram & Sommer, 2019). The scope of activities and the effects achieved in the process hardly differ in peacetime and wartime in this regard. At the other end of the theory spectrum is a war in which massive cyberattacks and the severe damage they cause to the enemy replace the use of conventional means of warfare and boots on the ground, or at least to a large extent replace the use of conventional means of action (Bigelow, 2019). Such an approach presupposes comprehensive planning and preparation, presumably lasting several years, as a component of military strategic planning (Buchanan & Cunningham, 2020), in which, after identifying relevant military IT targets in cyberspace, these are specifically infiltrated even in peacetime by means of complex hacking attacks and provided with the possibilities of military intervention (which, depending on the planning, can range from disruption to destruction) (Biller & Schmitt, 2019). This initial preparation is then followed by a phase of careful maintenance of this effective capability in order to conceal it from detection and the associated closure of access until a possible deployment (Reuter, Riebe, et al., 2019). With regard to relevant targets within such a cyberwar, there is concern, among other things, that in addition to attacks on military systems, IT systems from the area of so-called critical infrastructures in particular will be (Global Commission on the Stability of Cyberspace, 2021) can become targets of cyberattacks. These systems are generally easily accessible via the Internet and often comply with generally known industry standards, in contrast to highly specialized military IT systems, which are often deeply embedded in military networks and correspondingly well protected in cyberspace due to their inherent security relevance (Mulazzani & Sarcia, 2011). Cyberattacks against such critical infrastructures can therefore be well-prepared, while triggering the disruptive effect in almost every case would cause significant and large-scale impairments to a state and thus directly limit its military options for action (Sandholz et al., 2020). Moreover, even the infiltration of the IT systems in question must be assessed as a significant risk in peacetime (Murphy, 2019), as attackers, in addition to the intended installation of the cyber means of action, can also unintentionally compromise the systems and their flawless operation and trigger domino effects due to the interdependencies of critical infrastructures. Although critical infrastructures are supposed to be protected during acts of war due to their importance to civil society, this certainty has been significantly shaken with Russia's war and the cancellation of the norms of state dos and don'ts that have been in place to date. This development fuels concerns that cyberattacks could also be carried out against other highly sensitive IT systems, such as nuclear weapons IT systems and their warning and control systems, in order to disable them and rob them of their deterrent effect (Eggers, 2021). At the same time, in this area, even the attempt to carry out cyberattacks and the associated targeted exploitation of security gaps and vulnerabilities is fraught with high risks due to the criticality of these IT systems.

In terms of expectations for cyber activities by Russia, until the war in Ukraine, the most notable activities in cyberspace in recent years were pronounced by Russian intelligence services (Greenberg, 2020; Trevithick, 2019). In Germany, for example, the two 2015

hacks of the German Bundestag's internal communications network called Parlakom caused (Guarnieri, 2015) or the hacking attack on the German government IT network of 2017/2018 (Tanriverdi, 2018) attracted attention. At the same time, Russia has also been identified as the originator, especially in the context of so-called hybrid threats (Ehrhart, 2019) and disinformation campaigns, as well as suspected state direction and direct influence of civilian patriotic hacking groups (US Department of Justice, 2020). In addition, the military significance of cyberspace has also been emphasized in official strategies and statements[2] and corresponding capabilities have been built up, for example, within the framework of the so-called Sofacy Group (Herpig & Reinhold, 2018), which is assigned to the GRU military intelligence service, so that many analysts have recently assumed that Russia has comprehensive military cyber capabilities.

A final important point that many were not aware of in Russia's current war against Ukraine, and which will be discussed in more detail below, concerns the role of non-state cyber actors. Until now, expected scenarios of cyberconflicts have ostensibly assumed military as well as intelligence forces as the primary actors, and international debates have focused on them (see Broeders and Cristiano, 2020 for an example). Even though cyberspace has always been a playground for criminal as well as civil activist groups, they have so far been less prominent in the context of interstate conflicts and their consideration. Although cyber-activists have played a limited role in regional conflicts, such as in Ukraine after the Russian invasion of 2014 (Maurer, 2015), their activities have largely been concerned with creating a public sphere for their respective political concerns and protesting against activities perceived as unjust. However, there has been no explicit participation in the context of interstate armed conflicts and no explicit attribution to one side of the conflict parties.

7.2 Assessment of the Role of Military and Offensive Cyber Means of Action in Russia's War Against Ukraine

In contrast to the expectations described above, a comprehensive cyberwar, in which military action in cyberspace plays a decisive role, has so far not occurred in the run-up to or during Russia's war against Ukraine - as far as publicly available sources suggest. In some debates, this has been taken as an opportunity to question the significance of cyberweapons in military conflicts in general, with reference to earlier warnings of cyberwar (Burton & Christou, 2021). Against this background, the offensive cyber activities in Russia's war against Ukraine will be examined below and evaluated within the framework of the continuum described at the beginning.

7.2.1 *Russia's effective cyber startup phase*

On the one hand, targeted and effective cyberattacks took place, especially in the initial phase of the war, suggesting long preparation phases in some cases. In addition to the massive (but technically relatively easy to implement) DDoS attacks on Ukrainian government websites, there was the targeted cyberattack on the satellite operator KA-

[2] See the "Cyber Policy Portal" of The United Nations Institute for Disarmament Research (UNIDIR), https://cyberpolicyportal.org

Sat (Schulze, 2022). In this case, terminals for communication with the satellite network were rendered unusable by manipulated software, which could only have been possible with extensive, months-long planning, technical analysis, preparation of tailored malicious code, and the creation of access points to introduce the malicious code into the satellite network provider's IT system. At the same time, it must be assumed that due to the necessary analyses for the execution of the KA-Sat hack, it must have been clear in advance that the release of the malicious code would not only damage terminals in Ukraine and that the impact outside Ukraine was therefore deliberately accepted[3]. The two initial cyberattacks carried out in the initial phase of the attack presumably served the purpose of preventing the dissemination of information and disrupting military defense systems in Ukraine (Hoppenstedt, 2022). In addition, Russia thus succeeded in significantly limiting the communication capabilities of the Ukrainian military, which could only be compensated for by the rapid assistance of the Starlink network (Ballweber & Reinhold, 2022). The necessary amount of preparation, precise scheduling of cyberattacks, and coordination with conventional military activities can be taken as evidence of the existence of military cyber capabilities and corresponding capacities, as well as the will to use this area for military purposes.

7.2.2 *Inadequate connectivity planning for cyberspace and international support for Ukraine*

Previous analyses of the Russian attack are largely in agreement that Russia's military planning assumed a very short, fierce warfare and a swift subjugation of Ukraine (Schwirtz et al., 2022). Presumably, the primary objective was to occupy strategically important infrastructure and to target and destroy militarily relevant supply routes. In the context of this planning, cyber activities were presumably prepared primarily to cause targeted disruption in the information space and to gain military advantage. At the same time, in addition to the elaborate preparation required for their use, cyber means are at a disadvantage compared to kinetic means in terms of accuracy and certainty in predicting and executing the desired effects when it comes to open, rapid action that does not require secrecy. The effect of a missile on a building, for example, is much clearer to plan and deploy from a military perspective. For these reasons, Russia had probably planned few other cyber activities, relied on the use of conventional forces, and did not plan for further coordination with military cyber units (Batemann, 2022). Thus, with the change in the war situation and due to the onset of international support (Beecroft, 2022; Corera, 2022) for Ukraine, including in the area of IT security and resilience, it was likely much more difficult for Russia to provide additional cyber assets in a timely manner in conjunction with the increased vigilance in the area of IT security of Ukrainian IT systems and infrastructure.

[3] See the analysis of the attack preparation and execution, as for example, extensively presented here and provided with further information: https://cyberlaw.ccdcoe.org/wiki/Viasat_KA-SAT_attack_(2022)

7.2.3 *Dangers of uncontrolled spread of cyber activities*

Taking into account all the possibilities related to cyberattacks, instead of the aforementioned cautious strategic planning in the selection and use of cyber means of action, Russia could potentially have chosen a more massive approach in cyberspace as the conflict prolonged. For example, instead of targeting strategically relevant targets, it would be conceivable to instead deploy malware against any available or vulnerable targets that could be detected and reached within Ukrainian cyberspace. Another variation could have been to deploy malware that has an automated propagation mechanism and can spread within affected networks or beyond IT networks to achieve an area effect. However, such a crowbar approach is associated with a significantly higher risk of unintentional and uncontrollable spread of deployed cyber agents beyond Ukrainian IT systems, as careful pre-analysis of affected systems and tailoring of malicious code to a specific target system is hardly feasible in a meaningful way due to lack of time. While the effects in the case of the KA-Sat hack were presumably known, and the risks of escalation were weighed and accepted, the crowbar approach described would pose an incalculable threat to IT systems in other countries and possibly affect Russian information and communications technology (ICT) systems themselves, thus inadvertently endangering themselves. Moreover, such action would put Russia at risk of revealing its own available cyber assets and knowledge of security vulnerabilities without simultaneously triggering targeted effects and thus rendering this knowledge ineffective in the sense of one-shot weapons. As reports have shown (ICRC, 2022b), critical infrastructure in particular has been the focus of Russian attacks for several months - including by means of cyberattacks (Seals, 2022). Despite the terrible damage caused and the associated consequences for the civilian population, however, these are rather long hanging fruits, since cyberattacks on such IT systems can usually be carried out using already known security vulnerabilities and correspondingly already available malware (Microsoft, 2022).

7.2.4 *Dependence on local and civilian ICT infrastructures*

Another related issue concerns the communications infrastructures of the Russian invasion forces, which were allegedly heavily dependent on Ukrainian ITC infrastructures. According to reports, the technical equipment of the conventional forces was so inadequate that combatants had to rely on their own private IT equipment and telecommunications and data connections had to be transmitted via Ukrainian ITC providers (so-called roaming). Large-scale interference with these communications frequencies and infrastructures or the disruption of navigation services such as GPS/GLONASS would have meant a massive restriction of their own communications and navigation capabilities. At the same time, it was precisely because the ICT infrastructure continued to function that the Ukrainian government and civil society were able to provide unprecedented coverage of the events of the war.

7.2.5 *Unexpected activities of non-state actors*

In particular, non-state actors on both sides of the conflict have undertaken enormous and unexpected activities in cyberspace. On the one hand, in response to a call from President Selenskyj, Ukraine was able to establish a civilian cyber unit - the so-called "IT Army of Ukraine" (Paganini, 2022b) - which was integrated into military planning and also used for disruptive activities against Russia. Second, Russia has emerged in recent years as a launching point for some very powerful non-state cyber actors, some of whom had previously been more prominent in the criminal sphere and some of whom had joined Russia's cause at the beginning of the war. In addition, semi-organized hacker and activist groups such as Anonymous (Pitrelli, 2022) or the "Belarusian Cyber Partisans" (Cox, 2022) have joined the Ukrainian side and massively attacked Russian IT systems (Nair, 2022). Even if the attribution of individual actions to the groups is difficult and hardly verifiable sustainably, their effect within Russia is likely to have meant that institutional IT specialists - especially with regard to intelligence services and the military - were needed for the self-security of the country's IT infrastructures, the administration and large private-sector companies, and were not available for use against Ukrainian IT systems. In particular, the resources of Russian non-state cyber actors were effectively tied up by so-called non-state vs. non-state hacking activities, or in some cases even completely disabled by data breaches - i.e., stealing and publishing confidential information. At the same time, it can be assumed that the numerous hacking attacks and data breaches against Russian state institutions, companies with state involvement, or news portals and information systems contributed to undermining the restrictive information policy regarding the war by putting Russia under pressure to explain itself domestically on the one hand and forcing it to react accordingly with harsh means, e.g., by banning large, internationally operating social media platforms, on the other.

7.3 Conclusions

Based on these observations and assessments, the paper considers the possible conclusions that Russia in particular and political and military forces in general might draw with regard to the future development of military and intelligence activities in cyberspace. Based on these considerations, the resulting challenges for international debates to contain the damaging use of cyber means of action are identified and possible solutions are presented.

7.3.1 *Increasing Threat to Critical Infrastructure*

In view of the very effective use of cyber means in the initial phase of war, it is to be feared that ICT infrastructures in particular will become a primary target at the beginning of warlike conflicts in the future - even more so than in the past. On the one hand, due to their nature, such systems are strongly IT-based and networked, so that a use of cyber means of action against them can be prepared and executed very effectively. In addition, chain reactions through downstream systems that depend on functioning ITC infrastructures can be exploited and the spread of a damaging effect can be planned for.

In this way, the communications capabilities of an entire country could be significantly and almost simultaneously limited, which would be impossible to implement by kinetic means on such a broad scale and in such a short time. In the case of Ukraine, significant redundancy of IT infrastructures and a highly heterogeneous landscape of many smaller and regional civilian ITC providers proved effective in compensating for the initial damage and limitations caused by such aggression and restoring ITC. At the same time, this level of resilience would presumably have been virtually impossible without civilian international IT security support and the rapid deployment of backup solutions, as well as the provision of redundant ICT infrastructures by international commercial players such as Starlink. Against this backdrop in particular, however, it can also be attested that the operational support provided to Ukraine - including in the area of cybersecurity - can be considered successful international crisis management, and this despite the fact that there are no established formal or institutionalized structures or organizations for this purpose yet. have contributed to resilience in the case of Ukraine.

7.3.2 *Increase in the preparation of offensive cyber operations*

Given Russia's lack of tactical capabilities to act adequately militarily in cyberspace beyond the initial phase of the war, it is very likely that future conflicts will be characterized by a more direct integration of cyber capabilities into conventional warfare to enable or support conventional tactical maneuvers and to maintain military pressure in cyberspace. However, in order to create such options for action and to have effective means available over the entire duration of a military conflict, massive cyber operations carried out covertly in foreign IT systems are already necessary in peacetime. In the process, actors will acquire high-value ITC targets and build and maintain backdoors into them so that they can be used for an attack at any time. Since defending against an acute cyberattack by means of a so-called hack-back is not very effective (cf. e.g. (Reinhold & Schulze, 2017)), such an approach could also be chosen as part of a defensively oriented defense strategy. In this case, a state would prepare for an anticipated defense case by preemptively preparing cyberattacks on high-value IT systems of an anticipated aggressor in order to be able to exert military pressure or restrict the military options of the anticipated aggressor in the event of a conflict. Since such options for action must be prepared in any case, intelligence or military action in foreign IT systems is already very likely in peacetime, even more so than is already the case. Since such a scenario creates considerable incentives to withhold security vulnerabilities in IT hardware and software as crucial basic material of any cyber and hacking activities and not to report them to the manufacturer for remediation, rules and agreements are needed here to contain the inherent large-scale threat to civilian IT systems.

7.3.3 *Need for internationally binding norms for state behavior in cyberspace*

Against the backdrop of the above points and the fact that any unauthorized intervention in foreign IT systems jeopardizes their secure functioning, the international debates on the dos and don'ts of state action and, in particular, the limits of activities in foreign IT systems in peacetime, which have been conducted for years in various formats, are becoming even more explosive. On the one hand, there are already comprehensive

proposals and approaches that have been developed, for example, within the framework of the UN Group of Governmental Experts on Advancing responsible State behavior in cyberspace in the context of international security[4] (UNGGE, 2021) or on the basis of best practice recommendations of the Organization for Security and Cooperation in Europe (Security & Europe, 2016). For example, the most recent UN GGE consensus report emphasizes, among other things, the validity of state due diligence in cyberspace (the so-called due diligence principle) and calls for the protection of critical infrastructure (Schulze & Datzer, 2021). At the same time, however, there is a lack of enforceability of these rules as well as an internationally uniform and, above all, binding perspective on how these and other existing norms of international law can be practically applied to cyberspace and to what extent states are required to implement these norms in their national sphere of competence or can be held liable in the event of non-compliance. At the same time, in view of the current tense global political situation, it is likely to be difficult to motivate Russia or even China as important players to engage in joint talks on global IT security. One common ground for achieving this nonetheless would be to appeal to the development of measures or limits to guarantee national IT security, since any actors are dependent on the guarantee and maintenance of national IT systems even in the event of damaging activities in foreign IT systems. However, this argumentation is contrasted by developments in China and Russia to decouple themselves from global IT infrastructures, manufacturing and supply processes, and hardware and software products. The more these countries use their own national solutions, the less willing they are likely to be not to unduly impair global IT products and IT systems in the interests of self-protection.

7.3.4 *Containment of the activities of non-state actors and state due diligence obligations*

With regard to the role and activities of non-state actors, the considerable challenge arises as to how their actions can be contained in future conflict situations in order to be able to control conflict dynamics, prevent endangerment of the civilian population and avoid international proliferation of the actors involved in the war. In particular, semi-organized civilian actors are likely to play a role in future conflicts. In this context, they can relatively quickly become relevant actors in cyberspace[5], using digital disruptive actions, hack-and-leak attacks, or targeted defacements to circumvent the conflict parties' information policies, undermine national support for activities, and challenge narratives, as well as tie up resources for national IT security and countering disruptive actions (Paganini, 2022a; Pitrelli, 2022). Particularly in the case of markedly asymmetric conflicts that exhibit a clear friend-foe division, it can be assumed that conflict parties will also find it easy to mobilize national or international non-state cyber

[4] For an explanation of the working group, see https://www.un.org/disarmament/group-of-governmental-experts/

[5] This is all the more true since hacking does not, in principle, require any specific military skills or tools. When selecting targets, groups like Anonymous have often focused on IT systems of public administration or of business and media companies in the past. Since such targets are usually publicly accessible from the Internet due to their regular use and are based on conventional IT systems, they do not require any high-value target information that could only be obtained through intelligence activities. Moreover, since such activities are more concerned with signaling to a perceived adversary than with specific longer-term disruptions, the selection of targets is generally flexible and does not relate to a very specific IT system, as is the case with military strategic planning.

actors and integrate them into state structures. However, this raises the question of the extent to which civilian actors could be perceived as military combatants and what the consequences would be for the "ius in bello" in particular and the perspective on war in general if cyber-combatants presumed to be actively intervening in a conflict are completely decoupled from the territorial limitations of a warlike confrontation. State conflicts can also extend to the level of cyberattacks directly between non-state actors and in turn endanger civilian IT systems. Against this backdrop, it should be examined in what form and to what extent governmental possibilities for exchange with these groups can be established in order to establish point-of-contacts or other communication possibilities for the case of conflict.

7.3.5 *Russia's Future Role in Cyberspace*

In view of Russia's behavior in the coming years, the Russian government will presumably continue and further intensify its strategy of hybrid conflict management against democracies due to its looming international isolation. It is likely that both disinformation and so-called digital disruption will continue to be used below the threshold of an open conflict in order to discredit open and free social systems, to serve its own victim narrative and to bind allies more closely to it. In this context, cyberspace continues to be very suitable for acting and discrediting from behind the scenes as well as for false-flag operations due to the possibilities of concealing the origin of activities. At the same time, it is not necessary for the strategy of digital disruption to specifically attack high-value targets - which are usually also particularly well protected - but rather it is sufficient to draw from the abundance of IT systems, which are unfortunately still poorly protected and inadequately maintained, to identify targets that are vulnerable to attack. Particularly in rural Germany, critical infrastructures are often operated by small and medium-sized enterprises (SMEs) whose IT security measures may be inadequate due to human or financial resources. This situation may not be much better for European partners and requires close and timely cooperation in order to be able to warn partners as quickly as possible when incidents occur. With regard to military action in cyberspace, it is to be feared that, although Russia has actively made proposals in recent years within the UN framework for pacification of cyberspace and the rules and limits of state action in this domain, the currently displayed disregard for established norms could also be implemented in cyberspace with attacks on civilian (critical) infrastructure, among other things. This development also makes the consensus reached in 2021 in the Group of Governmental Experts on Cyber Issues (UN GGE) (UNODA, 2021), which was originally regarded as substantial progress, on the validity of existing cyber norms and, in particular, the state's responsibility for due diligence, appear unpromising.

7.4 Summary

In summary, although Russia's war against Ukraine was not accompanied by the expected massive use of military assets in cyberspace, this can only be attributed to a lesser military relevance of this domain to a limited extent. Rather, the unexpected role of cyberwarfare is more likely to be due to the specific situation, tactical planning failures and misjudgments by Russia, and in the further development of the war. At the same time,

it must be assumed that both the state actors involved and other states will learn their lessons from Russia's alleged tactical and strategic mistakes and adjust their strategies, activities, and military planning accordingly. In sum, this suggests that, with respect to the expectations explained at the outset, a full cyberwar as a boots-on-the-ground substitute is rather unlikely. At the same time, however, cyberspace will continue to gain in importance as a military domain in the future, especially with regard to tactical planning in the area of interdiction and disruption of communications and supply infrastructures. To the same extent, the intelligence-focused significance of cyberspace outlined above is thus also gaining relevance; this is particularly true in peacetime with regard to preparations for armed conflicts as well as in the provision and protection of national IT systems and the monitoring of foreign actors. These conclusions, and the extent to which it is possible to bring the parties to a conflict together for joint discussions despite increasing international bloc formation, and to include non-state actors where appropriate, will largely determine the shape and scope that cyberspace will play in future conflicts. This will also determine whether it will be possible to ensure the peaceful use and further development of this globally used domain and to achieve a containment of the current tendencies toward militarization of this domain.

8

ARMS CONTROL AND ITS APPLICABILITY TO CYBERSPACE

ABSTRACT Arms control aims at preventing conflicts and fostering stability in interstate relations by either reducing the probability of usage of a specific weapon or regulating its use and thus, reducing the costs of armament. Several approaches to arms control exist: limiting or reducing numbers of weapons and armed forces, disarmament ("down to zero") or prohibiting certain weapons. To illustrate these further, this chapter elaborates on the necessity of arms control and presents some historical examples, including an overview of existing measures of arms control. Extrapolating from these, the general architecture of arms control regimes and the complex issue of establishing and verifying compliance with agreements will be discussed, not least with respect to cyberspace. Building on these theoretical considerations, the chapter presents important treaties and first approaches, including the Wassenaar Arrangement, the recommendations of the OSCE, and the UN GGE 2015.

ORIGINAL PUBLICATION *Reinhold, T., & Reuter, C. (2019a, March 13). Arms Control and its Applicability to Cyberspace. In C. Reuter (Ed.),* Information Technology for Peace and Security: IT Applications and Infrastructures in Conflicts, Crises, War, and Peace *(pp. 207–231). Springer Vieweg. https://doi.org/10.1007/978-3-658-25652-4_10*

8.1 WHAT IS ARMS CONTROL AND WHY IS IT NECESSARY

The concept of arms control has been developed as a political reaction to the dynamics of military armaments in the international state system. At its core, **arms control** is a normative endeavor. It was born out of the insight that nuclear war needs to be prevented, and it is guided by the principle of preventing future wars. The concept can be described as "*unilateral measures, bilateral and multilateral agreements as well as informal regimes (. . .) between States to limit or reduce certain categories of weapons or military operations in order to achieve stable military balances and thus diminish tensions and the possibility of large-scale armed conflict*" (Den Dekker, 2004). Thus, arms control does not necessarily imply steering armed forces towards complete disarmament. Early attempts of arms control can be recorded in the pre-20th century, often accompanying larger conflicts or new military technologies like the development of firearms and high caliber guns. These early approaches, like The Hague Conventions of 1899 and 1907 and their annexes[1], often included the non-usage of certain weapons such as chemical weaponry. This dynamic increased with the advancements of military weapons during

[1]Both Hague Conventions from 1899 and 1907 consist of multiple treaties and additional annexes. Most relevant for the challenges of arms control is the second treaty of the first conference "Convention (II) with Respect to the Laws and Customs of War on Land and its annex: Regulations concerning the Laws and Customs of War on Land. The Hague, 29 July 1899" (The Hague Conference, 1899) as well as the fourth treaty of the second Hague convention (The Hague Conference, 1907).

the First and Second World War as well as with the subsequent arms races of the Cold War. Especially the development of nuclear weapons, their massive destructive potential and the high risk of global annihilation underlined the necessity of political regulation of these developments. Arms control is usually conducted in the form of bilateral or multilateral legally binding treaties to regulate some aspects of military potential and capabilities, but it is also concerned with the conditions and circumstances that lead to armed conflicts. The overall goal of arms control is less a complete disarmament which – strictly speaking – would mean the renunciation of all military capabilities, but rather a rational planning for reducing the risk of war. This task can be divided into three different parts (Müller & Schörnig, 2006):

1. War prevention and the reduction of conflict probability, limiting the acceleration of armament dynamics and its causes and reducing the likelihood of preventive or preemptive strikes.
2. Damage limitation in the event of armed conflicts, restricting the extent of death and destruction by certain weapon systems with massive destructive potential or weapons that can be used on a large scale.
3. Reduction of armament-related costs and the release of such funds.

Against the background of these overall tasks, arms control approaches generally consider the following different principles and measures specified in individual and usually legally binding treaties for specific weapons, weapon parts, weaponizable technologies, and armed forces:

- Create transparency about military capabilities, establish and maintain sustainable stability and communication in inter-state relations, so-called Confidence and Security Building Measures (CSBMs or CBMs).
- Provide quantitative and qualitative limits of allowed weapons or its specific capabilities, for instance, the payload or the range of missiles.
- Restrict or prohibit the proliferation of weapons, weapon parts or weapon technology, establish measures to control restrictions or limitations and provide information to other states about arms sales.
- Develop and establish specific measures of verification that enable states to practically verify the compliance of other treaty parties with agreements.

These approaches are not necessarily consistent or compatible, and the particular focus in a concrete situation as well as the corresponding means always depend on the configuration and level of political, economic or (expected) military conflict. This is also important in view of the realistic assessment of possibilities and expected results of arms control in specific situations and their limitations. Therefore, arms control cannot be equated with **disarmament**. This may be the case, for example, when limits are set for weapon systems that are above the current stock levels of two treaty parties. The controlled armament build-up to the new limits could allow a balance of military

power and reduce concerns of a later and possibly hidden armament. In general, arms control stretches from measures with minimal requirements of commitment to establish first steps for positive state relations to reduction measures with practical controls and monitoring of weapon sites or other relevant facilities. Figure 8.1 shows the "Non-Violence" sculpture in front of the UN headquarter – a classical tribute to non-violence and peace.

Figure 8.1: Sculpture "Non-violence" showing a revolver tied in a knot, on display outside the Headquarters of the United Nations in New York City by the sculptor Carl Fredrik Reuterswärd (Picture: C. Reuter)

8.1.1 *Historical Examples of Arms Control*

Some examples aim at illustrating that over the last decades, each new emerging military technology raised new challenges for arms control, led to international debates and – often after their military deployment – to agreements and treaties[2].

8.1.2 *Arms Control for Nuclear Weapons Technology*

Due to their major threat to mankind and the historical arms race during the Cold War era, the regulation of nuclear weapons and its carriers like missiles and warheads has a long history with many, sometimes unsuccessful, approaches of mutual agreements and

[2]For an insightful overview of arms control endeavors see Goldblat (Goldblat, 2002)

treaties. The following examples also illustrate a specific aspect of arms control treaties. In most cases, the agreements not only have a specific technological or military-strategic scope but also a limited period of validity. Often, they are intended to be reviewed and possibly renewed after some time or followed by subsequent treaties. Because of these expiration dates or the unilateral cancellation of treaty signatories, some of the agreements were terminated without follow-up approaches. The list further exemplifies that arms control regulation is often a step-by-step process, starting with minimum consensus regulations proceeding towards stricter prohibitions. This development can be seen in the first arms control agreement for nuclear weapons and weapons technology, the so-called Partial Nuclear Test Ban Treaty (PTBT)[3] which entered into force in 1963 (UN, 1963).

The treaty was initially signed by the Soviet Union, the United Kingdom, and the United States and then opened for signature by other countries. It prohibits all test detonations of nuclear weapons other than those conducted underground and is still active. The agreement can be perceived as a first measure to slow down the nuclear arms race and its proliferation by limiting the scientific testing capabilities. A few years later, in 1970, the Non-Proliferation Treaty (NPT)[4] came into force, taking arms control of nuclear weapons an important step further (NPT, 1970). The treaty is based on three pillars.

1. It firstly defines a list of nuclear-weapon states that have manufactured and exploded a nuclear weapon or other nuclear explosive devices before 1 January 1967 and declares that all other non-nuclear weapon states agree to never acquire nuclear weapons.
2. Its second pillar is the agreement of all treaty parties to pursue nuclear disarmament in order to ultimately eliminate nuclear arsenals (Campbell et al., 2005).
3. Its third pillar is the right of all parties to develop nuclear energy for peaceful purposes and to benefit from international cooperation in this area.

The NPT originally had a limited duration of 25 years but was extended indefinitely in May 1995. It is now reviewed every five years in the Review Conferences of the Parties. An important aspect of the NPT is that it authorizes the International Atomic Energy Agency (IAEA) to monitor the states' compliance with NPT agreements and commits them to security measures, the so-called safeguards.

Another issue of arms control is highlighted by the 1988 Intermediate-Range Nuclear Forces Treaty[5] between the United States and the Soviet Union (INF, 1988). The treaty does not focus on the nuclear explosive device itself, but on its deployment tools, the missiles and the necessary launchers. It codified the elimination of all nuclear and conventional missiles and their launchers with specific ranges and ordered a deadline for their destruction. In addition, verification measures such as on-site inspections were

[3]The full name of the treaty is "Treaty Banning Nuclear Weapon Tests in the Atmosphere, in Outer Space and Under Water", but it is also known as Limited Test Ban Treaty (LTBT).

[4]The full name of the treaty is "Treaty on the Non-Proliferation of Nuclear Weapons".

[5]The full name of the treaty is "Treaty Between the United States of America and the Union of Soviet Socialist Republics on the Elimination of Their Intermediate-Range and Shorter-Range Missiles".

established to check compliance with the treaty by both sides. Besides the obvious positive effect of reducing the military escalation potential of nuclear weapons, the agreed verification measures are valued by peace and security researchers because they established specific, practical and measurable steps[6] for checking compliance while respecting and sustaining national security agendas. After many years of criticism against Russia for undermining the agreements, the INF treaty is currently on the brink of termination as the United States has announced its withdrawal in February 2019. A similar fate of non-prolongation is threatening the so-called New START treaty that was signed in 2010 and entered into force in 2011 (NEW Start, 2010). START is the abbreviation for Strategic Arms Reduction Treaty and is used to describe three different, consecutive treaties between the Soviet Union (later Russia) and the United States on the reduction of nuclear bombers, intercontinental and submarine-launched ballistic missiles and warheads in combination with the establishment of verification measures. The New START treaty is expected to last at least until 2021, but negotiations for a follow-up treaty are currently not pursued.

8.1.3 *Arms Control for Biological and Chemical Weapons Technology*

As mentioned, arms control treaties were also negotiated for many other technologies. Two other important weapons of mass destruction are chemical or biological weapons. Facing the challenges and risks associated with them, the member states of the United Nations adopted the Biological and Toxin Weapons Convention (BWC)[7] that entered into force in 1975 which prohibits the development, production, stockpiling and distribution of biological weapons in combination with the strong emphasis on restricting the application of biological and toxic material to civil purposes (UN, 1972a). Since its implementation, review conferences have been held every five years. However, in the absence of specific compliance or verification stipulations in the treaty, effective monitoring of compliance has proved to be insufficient. Attempts to solve this problem by means of an additional protocol, including disclosure requirements and inspections, failed in 2001. As for the challenge of chemical weapons, the Chemical Weapons Convention (CWC)[8], signed in 1993 and entered into force in 1997, provides a series of comprehensive and practical disarmament steps (CWC, 1997). The signatory states undertake to declare existing stocks and to destroy all chemical weapons by 2012, a deadline that had to be prolonged, under international supervision. In addition to toxic chemicals, the CWC also applies to ammunition or equipment specifically designed to cause death or other harm by exploiting the toxic properties of the listed chemicals. The CWC also included the establishment and authorization of the Organization for the Prohibition of Chemical Weapons (OVCW), based in The Hague, which is responsible for monitoring compliance with the Convention. In a so-called "verification annex" to the Convention, contractual obligations (i.e. a detailed description of procedures to be followed by the treaty parties) and verification measures (i.e. how inspections are to be conducted and how samples are to be collected, handled and analyzed) are specified.

[6]Verification measures include extensive data exchange, on-site inspections at deployment sites, permanent inspections at the missile production facilities (Woolf, 2010).

[7]The full name of the treaty is "Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on their Destruction".

[8]The full name of the treaty is "Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction".

8.1.4 *Arms Control Treaties for Conventional Weapons and the Outer Space*

Another example for the diverse field of arms control approaches is the Outer Space Treaty[9] from 1967. Its aim is to prevent the occupation of celestial bodies by individual states (at that time: the Soviet Union and the USA) and the temporary or permanent deployment of military forces in space, on the moon or other celestial bodies, especially weapons of mass destruction (UN, 1967). However, given the spirit of technological advancement, civil space exploration is explicitly allowed for each state. Regarding arms control for conventional forces and weapons, the 1990 Treaty on Conventional Armed Forces in Europe (CFE) sets upper limits for the number of heavy weapons systems that may be deployed in Europe (CFE, 1990). After its implementation, the treaty led to drastic reductions in stocks of weapons that could be used for offensive purposes in Europe as a stable balance of military powers between the Cold War parties was established. One last example to mention is the Convention on Cluster Munitions (CCM, 2008). The CCM is a ban on the use, manufacture and transfer of certain types of conventional cluster munitions. It refers to bombs, grenades or warheads that do not explode as a whole but release a variety of smaller explosive devices. In addition to the prohibition provisions, the agreement includes provisions on the destruction of existing stocks, the disposal of residues from cluster munitions and the support of victims of cluster bombs. The convention was signed in December 2008.

8.1.5 *Preventive Arms Control*

One concept of arms control that is useful in assessing uncertain scenarios such as the militarization of cyberspace and the many technical difficulties associated with it is the so-called **preventive arms control**. It complements traditional arms control by focusing on technologies that are still in the research and development stages today. Preventive arms control attempts to regulate, limit or minimize technological innovations that could have negative effects on international security and peace to prevent such consequences as early as possible. The assessment of preventive arms control follows three main objectives (Neuneck & Mölling, C., 2001):

- Risk prevention for sustainable development and the evaluation of the consequences and potential dangers of the technology for the human, environmental, social and political systems and infrastructure complexes.
- The further development of effective arms control, disarmament and international law to place new technologies under existing arms control and disarmament contracts or existing international treaties as well as the development of new standards.
- The reduction or limitation of the extent to which technologies have destabilizing and negative effects on international security, either as

[9] The full name of the treaty is “Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies”.

a result of qualitative armament or in terms of the proliferation of armament-related knowledge.

8.1.6 *An Overview on Existing Measures of Arms Control*

An important step towards arms control measures regarding the militarization of cyberspace is to look at the history of similar measures of former technological developments and their military application. The specific requirements, technical constraints and goals of these approaches, as well as the lessons learned of their success or failure, are a valuable resource for their application to cyberspace. The following Table 8.1 depicts a categorized list of arms control measures (Neuneck & Mölling, C., 2001; Stohl & Grillot, 2009):ö

Forms of Arms Control	**Explanations and Examples**
Geographical measure	Demilitarised regions, security zones, e.g. nuclear weapon-free zone Africa
Structural measures	Defensive orientation of force structures, e.g. the Treaty on Conventional Armed Forces in Europe (CFE, 1990)
Operational measures	Limitation of manoeuvres, omission of provocative actions e.g. the Vienna Document (OSCE, 2011)
Verification measures	Data exchange, inspections etc., e.g. the Open Skies Treaty (OSCE, 1992) or the IAEA Nuclear Safeguards in Iran (IAEA, 2015a)
Declaratory measures	Waiver of the first use of weapons, especially nuclear weapons
Technology-/Medium-related measures	Limitation, reduction or destruction of certain weapons or technologies, e.g. ABM Treaty (UN, 1972b), INF Treaty (INF, 1988), individual marking of weapons to make the flow and illegal discharge of weapons comprehensible e.g. Arms Trade Treaty (UN, 2013)
Proliferation-related measures	Prohibition or restriction on the export of militarily relevant technologies, e.g. Nuclear Suppliers Group under the NPT (NPT, 1970), securing the storage and production facilities of weapons to prevent illegal diffusion
Application-related measures	Prohibition or restriction of the use of certain weapons and methods of war
Actor-related measures	Prohibition, restrictions or permissions in relation to specific groups of actors
Target-related measures	Safeguard clauses, prohibition of the attack on certain, especially civil, targets, e.g. the treaties of the Geneva Convention (ICRC, 1949)
Economic/Trade-related measures	Registration and licensing of arms dealers, producers, shippers as well as the regulation and approval of individual arms transfers and provision of sanctions and intervention options, licensing arrangements for import, export, transit through national territories of weapons

Forms of Arms Control	Explanations and Examples
Interstate cooperation measures	Inter-agency coordination, cooperation, coordination between relevant governmental organisations involved in arms control and, if necessary, cooperation in law enforcement with appropriate powers of the commissioned institutions
Information exchange measures	Transparency of production, ownership, trading and control efforts and dissemination of information to international partners

Table 8.1: Forms of arms control

8.2 Arms Control Measures

8.2.1 *Confidence Building and Verification as Important Parts of Arms Control Measures*

The historical examples showed that arms control efforts are almost always a gradual process; their success is often temporary and a matter of the political circumstances and responsible actors. In many cases, the initial situation is characterized by two or more state parties with a certain degree of mistrust or uncertainties about the current or planned military power and security policies of "the other sides". These situations, sometimes combined with ideological differences, have often been marked by little official communication. Each party depends on the "outside perception" of other parties and the interpretation of their actions, without having complete knowledge about their intentions and motivations. These constellations can be described by the sociological system theory of Parsons and Luhmann and their concept of "double contingency" (Luhmann, 1984). Applied to the context of international security politics, this means that state parties are under the impression of existing or perceived threats of other state actors that will or may interfere with their national security, sovereignty or foreign policy goals. Such threats can be an aggressive territorial behavior but also military armament which is perceived as overpowering either in terms of shear capacities of military power (e.g. conventional forces like tanks, infantry, military airplanes) or by the destructive military potential of specific weapons technology. Such tense situations are often exacerbated by new technologies and the inadequate or missing understanding of their invasive or destructive capacities.

The current debates on cyberweapons illustrate this situation: It is yet unclear what cyberweapons are and if cyber-related offensive military acts fit the conventional term of use of "**weapons**". As Sommer and Brown point out "*there is an important distinction between something that causes unpleasant or even deadly effects and a weapon*" (Sommer & Brown, 2011). The authors define: "*A weapon is 'directed force' – its release can be controlled, there is a reasonable forecast of the effects it will have, and it will not damage the user, his friends or innocent third parties*". Another approach for the definition of cyberweapons proposes an assessment of the strategic selection of the target, the purpose and the intended damage of specific cyber incidents and the attackers behind. Despite this rather terminological debate, there have been several interrupting

and sometimes damaging incidents in cyberspace. International studies emphasize the increasing demand of military forces for cyber-related capacities (UNIDIR, 2013). On the other hand, it is unclear how to measure, compare and categorize such cyber tools and their potential military destructive effects. As a result, especially in political debates, each state expects the most dystopian scenarios and tries to prepare for them, either with cyber defense measures or sometimes by setting up its own offensive cyber capacities. The ongoing debates about active cyber defense (in Germany known as the "Hack-Back" debates) or the perpetual fear that critical infrastructures could be shut down by military cyberattacks are the most visible parts of this Zeitgeist. However, there are no empirical studies that suggest the likelihood of such incidents.

In the face of these challenges, relations of mistrust, armament and the risk of conflicts by accident or misconception, the international political community has developed the concept of confidence building measures (CBMs)[10]. These measures, originally introduced by the Conference on Security and Co-operation in Europe (CSCE) during the Cold War era, intend to establish cooperation between states through gradual and mutual concessions, exchange of information and the reduction of military threats (OSCE, 1986). The proposed actions further intend to establish active channels of communication between opposing parties, facilitating communication in times of crisis before "pushing the buttons". The exchange of information and talks about national security doctrines or strategies and the underlying motivations aim at fostering an understanding of the security goals and fears of the "other side". At best, they could help the parties reach the common understanding that weapons should be seen as "military insurance" and not be used. Such situation emerged, for example, during the Cold War, where the capacities of nuclear weapons either reached a level that ensured a balance of power between the opposing states or provided the military tactical possibility for an immediate strike back[11]. Over the last decades and especially in times of the Cold War, some trust building approaches explicitly focused on technical-level talks about aspects of securing weapons and their facilities. Protecting one's own population from unwanted and destructive effects of weapon technologies by accidents can be seen as the least common denominator of all states.

These approaches sometimes helped to circumvent the ideological differences that would otherwise overshadow or even prevented these exchanges of knowledge. Such talks and conferences, more specifically, the establishment of mutual understanding, often became the starting point for further debates about reducing or stopping arms races. Moreover, they promoted agreements that kept a balanced level of specific weapons that sufficed for all sides in terms of their national security considerations without further armament. The fact that many of the above-mentioned examples of weapons technology also contain potential risks for the civil society and risks of technical accidents helped to drive debates further towards the reduction of military capacities or the proscription or the abolishment of specific weapon technologies.

As mentioned, the general goal of any arms control agreement or treaty is reducing the likelihood of war by a reduction of military technology weapons, their development,

[10] In debates that directly address military forces, the term is often extended to confidence and security building measures (CSBM)

[11] The military concept of a strike back followed the deterrence idea of preventing the threat of a nuclear attack by a country's assured ability to respond with an own nuclear attack. Such a "second strike" should have destroyed the attacker too, and by that minimised its intent for the first strike.

testing or military application. To restrict or regulate these aspects, treaties define rules for forbidden activities, thresholds for the numbers or instructions for the handling of specific items. The stability of arms control treaties depends on the widespread acceptance and support of these rules, as well as on the existence of trustworthy and effective compliance procedures (Müller & Schörnig, 2006). This underlines the importance of possibilities for treaty parties to check compliance with agreements of other parties, especially when the mutual relationship is characterized by mistrust. This vital part of arms control treaties can be implemented in different ways, and the agreed measures are specific to the regulated technological issues and the political goals of the negotiating parties. These so-called verification measures range from methods that allow supervision without on-site assessment like aerial imaging or seismic sensors to the structured collection, submission and exchange of data between states on stockpiles and trade volumes and on-site inspections with counting and measuring stockpiles and facilities. Müller & Schörnig (Müller & Schörnig, 2006) define four important characteristics for the acceptance of these measures:

- Appropriate and focused on the given context and the intended regulation of the selected items.
- Practicable and able to detect violations.
- Adequate and suitable to assess violations and their military dimension.
- Effective to recognize violations without being hindered by technical obstacles or political intentions.

8.2.2 *The Challenges of Arms Control Measures in Cyberspace*

Cyberspace as a domain has some very specific characteristics that are very different from other domains like land, air and sea. This includes the virtuality of this field and the information it contains, the non-physical representation of code and the seamless duplication of data. These features pose many challenges, especially for the practical side of arms control agreements; many of the established approaches will not work. In particular, this concerns all measures that rely on one of the following aspects:

- The limitation or the reduction of cyberweapons.
- The differentiation between civil and military usage and the resulting differences in authorization.
- The differentiation between a defensive and an offensive usage of cyber tools.
- The assignment of responsibility for individual activities in this domain.
- The necessity to practically control or monitor compliance with agreements.

Chapter 17 *"Verification in Cyberspace"* will have a detailed look at the specific technical aspects of cyberspace that cause these challenges and explain how cyberspace differs from real physical domains. The chapter will further explain how to deal with these problems and what aspects and measurable parameter could be used to implement verification measures for this space.

The previous examples of arms control approaches have shown that many of the approaches are based on states' declarations of the intended use or non-use as well as the trade or exchange of information on restricted items. Nevertheless, the ongoing international political debates currently struggle to find a way to reach binding agreements in the cyber area. Besides the technical difficulties preventing a one-to-one application of established measures to the new domain, this has many reasons. One of these problems is based on the different views of states about what constitutes cyberspace and the question of state sovereignty in this area. Whereas proposals from European states or the US usually focus on the IT infrastructure and acknowledge human rights and the freedom of speech, other approaches, such as a proposal to the UN by Russia, China and other states (UN General Assembly, 2011), emphasize the national right to monitor and regulate the distribution of information in this space. This potentially includes censorship. This conceptual disagreement is further complicated by the problem of transferring the idea of national borders to this area; determining the sovereign territory of a state as well as the area of its responsibility is complex. Another aspect exacerbating these disagreements is the question of which international committee or institution can be entrusted with monitoring and controlling the further technological development of cyberspace supporting its long-term peaceful orientation. This task was historically taken by different organizations like the Internet Corporation for Assigned Teams and Numbers (ICANN) and the Internet Engineering Task Force (IETF), which did not represent the international state community and may have been influenced by individual state actors. Approaches to transferring these tasks to a UN institution such as the International Telecommunication Union (ITU) have so far been unsuccessful. A similar question arises regarding an internationally legitimate institution that could be assigned with investigating suspected state-actor-driven incidents that would require (in most cases) the exchange and analysis of malware samples or sensitive log data from the affected IT systems (Davis II et al., 2017). A further problem for arms control approaches is the current lack of internationally consistent classification of cyberweapons or any kind of malicious cyber tools, such as exploits and vulnerabilities in IT products. This lack prevents a uniform risk assessment. Thus, there is no basis for any kind of definition specifying limitations or reporting obligations. This applies in particular to the necessary analysis of possible damage and the classification of different types, ranges and destructive factors of cyberweapons. The lack of classification further intensifies cyber armament as unpredictability hinders a "stable balance of military cyber power" where states would agree to limit military capabilities that meet their security requirements.

Previous cyber incidents showed that cyberweapons have so far - unlike expected – mostly been used for gaining hidden accesses to IT systems. This resembles espionage tactics rather than the use of classic weapons with disruptive or destructive effects. Cyberweapons rely in most cases on the exploitation of vulnerabilities in IT products. Thus, they are "one-shot weapons" that lose their impact once released because they reveal their attack vector and the exploited weakness in other systems can be closed. This results in a very cautious disclosure of the cyber capacities of states which - from

a military tactical perspective - work best when they are secretly implanted into the targeted systems and stay hidden until their application is needed (OSCE, 1992).

8.3 Important First Approaches of Arms Control in Cyberspace

As demonstrated, there is a growing international understanding of the dangers of an uncontrolled militarization of cyberspace and the need for cyber arms control measures. The historical examples illustrated that the first step for specific agreements on the limitation or the reduction of military goods is a common understanding of the problems and the risks of the technology. The debates within the international community are moving in this direction, forming an essential basis for agreements on norms and rules for state behavior in cyberspace as well as for binding future treaties on the military usage of cyberspace technology. The last part of this chapter will present some of the attempts that have been made in recent years by various actors and at different levels of inter-state cooperation that have driven these debates forward and will hopefully help pave the way towards broader agreements. The approaches are not ordered chronologically but according to the involved stakeholders and their target group. It is important to mention that these examples do not always explicitly fulfil the criteria of arms control treaties in accordance with the presented historical treaties and agreements. Their selection will present state-driven initiatives as well as proposals from economic actors and the civil society to illustrate the different aspects of the ongoing debates in cyberspace and their challenges, as well as the first results of these efforts.

8.3.1 *The Wassenaar Export Control Arrangement and its Extension from 2013*

The "Wassenaar Arrangement on Export Controls of Conventional Weapons and Dual-Use Goods and Technologies" is a multilateral export control regime. It was established in 1996 and currently consists of 42 member states (The Wassenaar Arrangement Secretariat, 2019). The objective of the Convention is to increase international transparency and regulation of trade, as well as to limit the distribution of conventional arms. The list of regulated items especially concerns so called dual-use items that can be used for both civil and military purposes. The member states of the arrangement undertake to control the export of these critical goods, examine export inquiries and, in the event of suspicion, reject them because of the potential for security-critical or human rights-endangering application. Trade data is exchanged between the member states twice a year. In view of the increasing expansion of intelligence and military activities into cyberspace, a first step towards regulating these activities was taken at the end of 2013. The extension of the agreement comprised the inclusion of "intrusion software" in the catalog of critical goods, regulated by the following definition (The Wassenaar Arrangement Secretariat, 2022):

> *" "Software" specially designed or modified to avoid detection by 'monitoring tools', or to defeat 'protective countermeasures', of a computer or network capable device, and performing any of the following: (a) The extraction of data or information, from a computer or network capable device, or the modification of system or user data; or*

(b) The modification of the standard execution path of a program or process in order to allow the execution of externally provided instructions."

This definition considers the functional scope of an application as a sufficient criterion for its regulation, less the possible damage or the specific application environment. One of the problems of the Wassenaar Arrangement is its implementation, which falls under the sovereignty and responsibility of each member state and is decided independently. In Germany, the Federal Office of Economics and Export Control (BAFA) has been commissioned to examine export inquiries. The German control criteria differ with regard to the destination of planned exports. Exports to EU Member States, NATO countries or states with a similar status are generally authorized unless there are specific political reasons against them. Exports to other countries are questioned and examined regarding the potential buyer, the possible open and hidden purpose of use, as well as the political situation and stability in the target country. These decisions and export controls are handled differently in other member states, and there is no obligation for standardized procedures. Control of the proliferation of such goods, an essential component of classical arms control agreements, is, therefore, only possible to a limited extent and does not achieve a universal validity. The approach could, therefore, be seen as a blueprint for a potentially global approach to regulating these goods and items if combined with consistent and equal national trade export laws and placed under an international control body such as a UN organization.

8.3.2 *The 2018 Proposal of the EU Parliament for a Harmonised Dual-Use Export Controls Regulation*

On the basis of the Wassenaar Arrangement, the European Commission has begun to discuss further regulation of such goods within the framework of a uniform export control system for EU countries (EU-Commission, 2016a). It prepared a proposal for the European Parliament, which adopted this position and prepared negotiations with the Council of the EU for a final agreement (Parliament, 2018). The EU Parliament's position follows most of the principles of the Wassenaar Arrangement on the regulation of technologies capable of cyber-surveillance and human rights violations. The definition of the proposal covers (EU-Commission, 2016b):

"items specially designed to enable the covert intrusion into information and telecommunication systems with a view to monitoring, extracting, collecting and analyzing data and/or incapacitating or damaging the targeted system. This includes items related to the following technology and equipment: (a) mobile telecommunication interception, equipment; (b) intrusion software; (c) monitoring centers; (d) lawful interception systems and data retention systems; (e) digital forensics;"

When assessing the export authorization for cyber-surveillance and other affected items, member states must consider the risk of infringement of the defined rules. This regulation potentially broadens the scope of regulated goods and their assessment in comparison to Wassenaar because it introduces a "catch-all control" approach which aims at supplementing the specific control categories for non-listed technology items and preparing regulation for future developments. Beyond the approach of an EU-wide common export control law, it also proposes a due diligence regime for exporting states and the exporter

itself, as well as a responsibility for standardized reports on national export control measures. This exceeds the Wassenaar approach of national sovereignty concerning the specific export rules and reporting procedures. In addition, member states may prohibit or impose an authorization requirement on the export of dual-use items not listed in the regulation for reasons of public security, human rights considerations or the prevention of acts of terrorism. The proposal of the EU Parliament is currently being discussed with the Council of the EU.

8.3.3 *Recommendations of the United Nations Group of Governmental Experts from 2015*

In 1999, the United Nations General Assembly passed the resolution 53/70 “Developments in the Field of Information and Telecommunications in the Context of International Security” (UN, 1999). The resolution is concerned with the increasingly relevant topic of cyberspace in terms of its potential for scientific and technological progress as well as its use for malicious purposes. A further resolution 58/32 of 2003 (UN, 2003) proposed to focus on the threats for this domain, the chances and possibilities for international cooperation in the field of information and communications technology (ICT) (including technical infrastructures) and established a group of governmental experts (GGE) to address these issues. Since its foundation, there have been five groups of governmental experts that were concerned with these questions and with the applicability of international law in cyberspace. Also, they prepared recommendations for international agreements. The last successful group from 2015 "*examined existing and potential threats arising from the use of ICTs by States"* and has recommended a set of voluntary, non-binding norms of responsible state behavior (UNGGE, 2015a). These norms have been adopted by the UN General Assembly "*in a call to its member states to be guided in their use of information and communications technologies. (..) G20 has also invited states to implement the GGE recommendations*" (UNODA, 2017). With regard to the challenges of arms control in cyberspace, the recommendations of the 2015 report addressed the following aspects:

> *“[It] recommended that States cooperate to prevent harmful ICT practices and should not knowingly allow their territory to be used for internationally wrongful acts using ICT. It called for the increased exchange of information and assistance to prosecute terrorist and criminal use of ICTs. (..) A State should not conduct or knowingly support ICT activity that intentionally damages or otherwise impairs the use and operation of critical infrastructure (..) States should not harm the information systems of the authorized emergency response teams of another State or use those teams to engage in malicious international activity. (..) States should take reasonable steps to ensure the integrity of the supply chain and prevent the proliferation of malicious ICT tools, techniques or harmful hidden functions. (..) The Group identified a number of voluntary confidence-building measures to increase transparency (..) and called for regular dialogue with broad participation under the auspices of the United Nations and through bilateral, regional and multilateral forums. (..) The report called for the international community to assist in improving the security of critical ICT infrastructure, help to develop technical skills and advise on appropriate legislation, strategies and regulation.” (UNGGE, 2015a)*

The 2016/2017 follow-up group did not reach a final consensus. This can be explained (among other things) by disagreements between states about the assessment of cyber incidents and their impact on national security. The members of the expert group could not agree on the question of how international law applies to the possibilities and limits of responses to such presumed state activities and appropriate countermeasures.

8.3.4 *Proposals for Confidence Building Measures by the OSCE*

Over the last years, the Organization for Security and Co-operation in Europe (OSCE) has issued two decisions that concern "confidence-building measures to reduce the risks of conflict stemming from the use of information and communication technologies". Decisions No. 1106 of 2013 (OSCE, 2013) and No. 1202 of 2016 (Security & Europe, 2016) are based on the organization's belief and commitment to foster international security by promoting communication and international cooperation between states and other relevant international organizations. In this regard, the organization developed a set of confidence building measures that should "*enhance interstate co-operation, transparency, predictability, and stability, and (..) reduce the risks of misperception, escalation, and conflict that may stem from the use of ICTs.*" The measures are voluntary, but the OSCE instructed its member states to base their political decisions, law-making and behavior on these principles. Most measures concern interstate consultations, the definition of a common terminology for cyberspace and its threats, the exchange of information regarding the security and use of ICTs as well as – in particular – the risks for critical national and international ICT infrastructures and their integrity:

"Participating States will nominate a contact point to facilitate pertinent communications and dialogue on security of and in the use of ICTs. Participating States will voluntarily provide contact data of existing official national structures that manage ICT-related incidents and co-ordinate responses to enable a direct dialogue and to facilitate interaction among responsible national bodies and experts. Participating States will update contact information annually and notify changes no later than thirty days after a change has occurred. Participating States will voluntarily establish measures to ensure rapid communication at policy levels of authority, to permit concerns to be raised at the national security level." (Security & Europe, 2016)

Furthermore, the proposal encourages the establishment of a central platform for the dialogue, exchange of best practices, awareness-raising and information on capacity-building as well as the handling of security threats and incidents and the OSCE is calling on its member states to prepare an effective national legislation for cooperation on this international, interstate level. The proposal also extended the considerations, especially regarding the significance of ICT for critical infrastructures and industrial IT systems, and encouraged its member states to cooperate in the exchange of national ICT incidents and the vulnerabilities detected. Although all these proposals concern "only" the political behavior of states (not the preparations of their armed forces) and are based on the exchange of information and the establishment of communication channels, these efforts must be considered as highly valuable. This is due to the important role of the OSCE as an international organization that connects states by providing an important and established platform for dialogue and decision-making, potentially fostering necessary

discussions and the finding of shared views and rules which could form a basis for negotiations and further agreements.

8.3.5 *State Driven Proposals for Norms and Responsibilities of State Behavior in Cyberspace*

Besides the previous multi-lateral approaches, various states have in recent years developed proposals for binding norms and rules of state behavior in cyberspace that followed established rules of international law. These proposals are often driven by national foreign policy priorities or reflect national views and concerns about state sovereignty and internal security.

At the end of October 2018, both Russia and the US, together with other supporting states, submitted two different proposals to the United Nations General Assembly First Committee for the further development of norms and responsibilities of state behavior in cyberspace. Both proposals assume that states should not use information technology to "carry out activities that are contrary to the maintenance of international peace and security" or "intervene in the internal affairs of other states". The Russian proposal (UN, 2018b), which is supported by 26 other countries, including China, reaffirms the UN GGE's recommendations. In doing so, the authors endorse a comprehensive list of international rules, norms and principles of responsible behavior. In particular, this draft resolution calls on the Secretary-General to convene an "open working group" to continue work on these issues which was discontinued by the UN GGE in 2017. A special feature of this proposal is, that it emphasizes the state sovereignty over the national internet in terms of the state's rights to examine and regulate the information that is shared, transferred, stored and distributed within national IT systems and the national part of the internet. The US-led proposal (UN, 2018a), supported by 35 nations, also confirms the UN GGE's work and calls for a further group of experts. In particular, it should focus on the question of how international law can be applied to the state's use of information and communication without defining new spaces of national sovereignty that profoundly conflict with freedom of speech and other human rights.

Two other proposals worth mentioning are the Paris Declaration and the Commonwealth Cyber Declaration, both published in 2018. The Paris Declaration was presented by the French government at the Internet Governance Forum (IGF) under the name of "Paris Call for Trust and Security in Cyberspace" (French Ministry for Europe and Foreign Affairs, 2018). The Call is formulated as a non-binding document and does not contain any detailed measures, nor does it propose to create new institutions. Rather, it aims to promote existing institutional mechanisms to "limit hacking and destabilizing activities" in cyberspace. This move intended to end the confrontations in the intergovernmental debates and the resulting stalemate. For this purpose, the call proposes that the monitoring of the effective implementation be delegated to the IGF as a UN body. The text contains nine objectives that balance its priorities between states, businesses and civil society, addressing three main issues: regulation of state-based activities based on norms, state sovereignty in cyberspace and protection of citizens. The document encourages more comprehensive and coordinated regulation of cyberspace, in particular, the maintenance of international peace and security. It not only recognizes the applicability of international humanitarian law to cyberspace, as well as international human rights law and

customary international law. The role and responsibilities of state actors in cyberconflicts are to be strengthened, and active cyber defensive measures by companies are excluded. In the same way, "offensive operations by non-state actors" and the influence of foreign states on democratic processes, such as elections, are condemned. Another central theme of the document is the importance of protecting individuals and critical infrastructures from harm. The document calls for the "public core of the Internet" to be protected from hostile actors and demands from the industry a stronger commitment to "security by design" in products and services. At the time of publication, the call was signed by 57 states, including the EU member states as the strongest faction. Russia, China and the US are not among the signatories.

A second declaration that is promoting similar goals is the "Commonwealth Cyber Declaration" (Commonwealth Secretariat, 2018) which was adopted at the 2018 meetings of the "Commonwealth Heads of Government Meeting". This is relevant in view of the many smaller and economically weaker states of this group, which emphasize the importance of cyberspace for their nations and express a right to co-determination in its development. The "Commonwealth Cyber Declaration" is, therefore, together with the OSCE CBMs, one of the strongest intergovernmental signals for the peaceful development of cyberspace so far. It acknowledges cyberspace as the basis of social, economic and political development and stresses the dangers of a destabilization of cyberspace by offensive state activities:

"We, as Commonwealth Heads of Government (..) recognizing the threats to stability in cyberspace and integrity of the critical infrastructure and affirming our shared commitment to fully abide by the principles and purposes of the Charter of the United Nations to mitigate these risks (..) commit to (..) limit the circumstances in which communication networks may be intentionally disrupted, consistent with applicable international and domestic law. We, as Commonwealth Heads of Government (..) recognize that without cybersecurity citizens are at risk of crime or exploitation, and commit to strengthening legislative, social and educational measures that protect the vulnerable." (Commonwealth Secretariat, 2018)

In this view, the declaration recognizes the importance of international cooperation in tackling cybercrime and promoting stability in cyberspace and supports the UN GGE's recommendations to develop frameworks for the application of international law to and establish confidence building measures for this domain.

8.4 Summary

The previous examples of international and national approaches to the development of binding rules and norms for state behavior have highlighted the increasing acceptance of the importance of cyberspace and the growing commitment of the international community to ensuring its stability. However, assessments, such as the 2013 Cybersecurity Index (UNIDIR, 2013), can only be the first step towards binding rules that limit, reduce or even prohibit the development, proliferation and usage of offensive cyber tools for military purposes. Besides the political will of states, many technical issues need to be analyzed to develop solutions to these challenges. Measures need to be developed that allow verifying compliance of treaty parties, the practical monitoring of military

facilities or the tracking of cyberweapon material like software vulnerability exploits. The history of arms control shows that this is a long way to go, but a necessary step towards the peaceful development of a global domain. To summarize the chapter:

- Arms control aims at preventing conflicts and fostering stability in interstate relations by either reducing the probability of the usage of a specific weapon or regulating its use and thus reducing the costs of armament. Thus, the overall goal of arms control is less a complete disarmament but a rational planning for reducing the risk of war.
- The field of arms control approaches is highly diverse; weapons to be controlled include nuclear, biological, chemical and conventional weaponry.
- Arms control measures include confidence building and verification or preventive measures.
- Cyberspace as a relatively new domain poses – due to its specific characteristics – many challenges. These include conceptual disagreements, the determination of territory and responsibility, as well as the establishment of a supervising authority. Many of the established approaches do not work.
- First approaches for a regulation of cyberweapons include the Wassenaar Export Control Arrangement and the 2018 Proposal of the EU Parliament for a Harmonized Dual-Use Export Control Regulation that could help to establish arms control measures in cyberspace.

9

CHALLENGES FOR CYBER ARMS CONTROL: A QUALITATIVE EXPERT INTERVIEW STUDY

ABSTRACT Cyberspace and its ongoing militarization have already been a topic in international fora as well as scientific debates for several years. Many important suggestions have been made for its regulation and development towards a peaceful application. However, the development of applicable, comprehensible, and verifiable arms control measures that can effectively reduce the risk of military escalations in cyberspace is still hindered by the characteristics of this domain. This article analyses these challenges and the obstacles of dual-use, proliferation, constant technological progress, the importance of the private sector, difficulties in defining and verifying the weapon, and difficulties in attributing attacks. By employing a literature review as well as qualitative expert interviews, the article provides a state-of-the-art perspective on these topics and answers the question to what extent expert knowledge aligns with the challenges discussed in the literature . It reveals the main challenges for cyber arms control - the lack of political will, of definitions, and of verification capabilities, as well as the dual-use nature of cyberspace and the multitude of stakeholders involved beyond states and discusses. Based on these findings, the analysis comes to the conclusion, that a broad definitional approach is advisable for cyberspace that regulates behaviors and outcomes rather than the technology itself. The analysis also shows that a reliable attribution mechanism is necessary for such endeavor, and that - as technical challenges have thus already been successfully overcome - an opportunity arises to translate the processes initiated under the UN GGE and the OEWG into binding regulations.

ORIGINAL PUBLICATION *Reinhold, T., Pleil, H., & Reuter, C. (2023). Challenges for Cyber Arms Control: A Qualitative Expert Interview Study*. Zeitschrift für Außen- und Sicherheitspolitik (ZfAS). *https://doi.org/10.1007/s12399-023-00960-w*

9.1 INTRODUCTION

In recent years, much has been written about cyberspace, its increasing militarization and its relevance in conflicts and combat. The still ongoing war in Ukraine presents a concrete case of an open military conflict that includes this digital domain. It also showed that despite the popular scholarly perception of a cyberwar that instantly brings down a state to its knees, military attacks in cyberspace can easily fail or be miscalculated and that a strongly decentralized IT infrastructure can withstand attempts to destroy a country's IT capabilities. While there is still much to analyze and learn from the war in Ukraine and the role of cyberspace in it, and even though it is very likely that military actors will draw their conclusions from mistakes and failures, one thing has become very clear: Cyberspace, beyond its relevance for civilian and commercial purposes, is also a military

domain. However, it is strongly influenced by factors that exceed traditional national jurisdiction and military power, such as by non-state actors entering a conflict, attacking each other's IT systems, or even fighting each other in cyberspace, or international commercial actors providing support for IT and communications infrastructures. A look at the past shows that arms control has been a successful means to respond to the security challenges of armament processes regarding different kinds of weapons and to contain or stop arms races, hence, significantly contributing to security and stability worldwide (Reinhold & Reuter, 2019a). Curbing the proliferation of weapons, as well as reducing those already in existence, is therefore one of the most important concerns of the international community (Müller, 2005). Despite the urgency emphasized by experts, like for example (Denning, 2001; Dittrich & Boening, 2017; Dunn, 2005; Hansel & Silomon, 2021; Litwak & King, 2015; Maybaum & Tölle, 2016), little progress has been made regarding arms control measures in cyberspace. Although increasingly established, at least in the normative field, so far these are mere declarations of intent or possible diplomatic measures in response to a cyberattack. However, approaches that attempt to address the complex matter of IT hard- and software, dual-use, and the question of what constitutes a cyberweapon (Reinhold & Reuter, 2021) are diverse, but so far lack a common international agreed definition. This article focuses on the debate regarding the challenges faced in establishing arms control in the domain of cyberspace, which include, e.g., dual-use, proliferation, constant technological progress, the importance of the private sector, difficulties in defining and verifying the weapon, and difficulties in attributing attacks. By employing a literature review as well as qualitative expert interviews, this study specifically aims to answer the following research question:

What are the current challenges for establishing arms control for cyberspace according to expert knowledge and to what extent do these align with the challenges discussed in the literature?

Since arms control represents an important instrument of foreign and security policy, this article aims to contribute to the interdisciplinary discourse on how international security in cyberspace could be maximized. Ruhmann emphasizes that "reality today requires new ways of thinking as well as recourse to usable known approaches" (translated by the author) (Ruhmann, 2010). Following this idea, this article adds to the ongoing discourse on the research and development of cyber arms control. Based on related work (chapter 9.2), the methodology used in this paper is presented in chapter 9.3. Subsequently, in chapter 9.4 the results of the expert interviews are presented regarding their perception of the challenges for arms control in cyberspace. The following discussion (chapter 9.5) is the core of this work and aims to answer the second part of the research question, to what extent expert knowledge and literature align regarding the challenges for arms control in cyberspace. Further, recommendations for academia and policymakers are presented before identifying the limitations of this study and considering starting points for future research. Finally, a conclusion is drawn in chapter 9.6.

9.2 Theoretical Perspective: Related Work

The present absence of treaty-based arms control in cyberspace can be explained by a number of factors, which will be discussed further in the following. This overview summarizes the current state of the art and explains why establishing arms control in

cyberspace is very difficult for several reasons. Figure 9.2 in the appendix provides a graphical representation of these findings.

A: Lack of a definition for the term cyberweapon: A fundamental challenge for establishing arms control in cyberspace is the lack of clear, uniform definitions of key terms, such as cyberweapon (Litwak & King, 2015; Reinhold, 2019c). The conventional definition of a weapon does not apply with respect to a cyberweapon (Arimatsu, 2012; Czosseck & Podins, 2012). Traditionally, a weapon describes an instrument of offensive or defensive combat, that is, a "device designed to kill, injure, disable or temporarily incapacitate people or destroy, damage, disable or temporarily incapacitate property or material" (US Air Force Secretary, 2018). While this definition of a weapon is applicable to kinetic weapons, it fails to capture the essence of a cyberweapon, which, unlike kinetic weapons, is in most cases not designed to produce a kinetic result that could possibly lead to one of the outcomes described in this definition (US Air Force Secretary, 2018). Hence, a cyberweapon becomes a weapon only through the attack capability of a malicious code in combination with a specific vulnerability within an IT software or hardware product as well as the intended result or effect of a code (Arimatsu, 2012). Much more decisive in this context is therefore for what purpose and with what intention such tools are used (Reinhold, 2019c). Additionally, it is not possible to differentiate between offensive and defensive weapons in the case of cyberweapons (Reinhold & Reuter, 2019a). Another perspective, suggested for example in the Tallinn Manual (CCDCOE, 2013), is to compare the actual effect triggered with the effects and impacts of physical weapons towards the loss of life or significant damage to objects. However, this approach is not suitable to classify and regulate malicious code prior to its actual usage. Due to the lack of clear definitions, the specificity required for legal regulations is lacking and significantly complicates the discussion of the topic (Arimatsu, 2012; Lewis, 2010a), especially as it is unclear what part of the malicious code should be regulated; either the knowledge and illicit trade with such knowledge, the code itself that exploits such knowledge to break into IT systems and circumvent IT security measures, or the actual payload which triggers the impact. Although a few approaches exist that try to focus on technical aspects of hard- and software to define a cyberweapon (Reinhold & Reuter, 2021), so far, no internationally accepted perspective exists, and thus it is not clear what should be discussed or negotiated at all (Geers, 2010).

B: Dual-Use-Dilemma: Another aspect complicating the search for appropriate cyberspace arms control is the dual-use factor by which cyberweapons are characterized (Reinhold, 2019c; Riebe & Reuter, 2019a). For example, a computer, a USB stick, or software can be used for both civilian and military purposes (Lewis, 2010a; Meyer, 2011) and even the knowledge about a security vulnerability is already an essential part of code that can be used to improve IT security as well as for an unauthorized intrusion into third-party IT systems. Therefore, no clear line can be drawn between these different use scenarios, which is why the products cannot be banned in principle in the context of arms control (Reinhold, 2019c). While dual-use has played a role in arms control treaties in the past, the dual-use nature of cyberweapons takes on a completely different dimension, as cyberweapons can be used in various ways (e.g., to destroy, degrade, exploit, control, deceive, or alter a target object). Therefore, the term dual-use can be misleading (Arimatsu, 2012; Reinhold, 2019c). Moreover, many instruments that could potentially be used as cyberweapons are also instruments for building a cyber defense or cyberespionage (Reinhold, 2019c).

C: Verification: Verification of arms control measures is one of the most central challenges for arms control in cyberspace (Arimatsu, 2012; Denning, 2001; Dunn, 2005; Lewis, 2010b; Maybaum & Tölle, 2016). One aspect concerns the previously described dual-use nature of cyberweapons: It is impossible to control the basic materials for building a cyberweapon because the technologies required to do so are either commercial or can be easily derived from widely available commercial products. Thus, it is not possible to distinguish the intended purpose of these systems based on their technical characteristics (Lewis, 2010a). Furthermore, cyber armament is happening covertly (Altmann, 2019b). Unlike, for example, tanks or missiles, cyberweapons are not visible as physical objects and can easily – or even must – be kept secret. Further, they are globally available, which means that it is not possible to limit weapons or capabilities numerically or spatially, as has been the case with weapons systems in the past (Altmann, 2019b; Dunn, 2005; Reinhold, 2019c). Additionally, due to their characteristics, cyberweapons (almost) cannot be detected by inspection teams or technical sensors (Maybaum & Tölle, 2016). And even if a cyberweapon was discovered, it would be impossible to eliminate all copies of it because cyberweapons or their components can be duplicated very quickly and inexpensively without the need for physical materials or special operating facilities (Altmann, 2019b; Dunn, 2005; Libicki, 2009), thereby preventing its non-proliferation. Hence, cyberweapons can be stored on computers and hard drives all over the world, e.g., even in locations under the jurisdiction of states that are not a party to any arms control treaty for cyberweapons and thus serve as "safe havens" (Dunn, 2005). Consequently, a geographical assignment of the data itself, or where it is stored or further processed, as well as the associated assignment to a specific national sovereignty and jurisdiction is difficult (Reinhold, 2019c). Considering these aspects, verification of malicious tools that are used in the cyberspace would require an extremely high level of interference, to which few if any states would agree (Denning, 2001). Moreover, it could be difficult because of the risk that states will be reluctant to share information about their capabilities in cyberspace, given the blurred line between capabilities that can be used as cyberweapons and those used for cyberespionage, which is generally not considered an act of war (Lewis, 2010b). In addition, intelligence agencies can use the same cyberweapons as armed forces, hackers, or criminals (Altmann, 2019b; Arimatsu, 2012; Czosseck & Podins, 2012; Libicki, 2009; Meyer, 2011).

D: Further technological progress and role of private sector: Adding to the previous challenges, tools for cyberattacks are changing very rapidly, which in turn makes monitoring compliance with treaties difficult (Denning, 2001; Dunn, 2005). This results in the fact that the development of new weapons and technologies, such as cyberweapons, have outpaced regulatory efforts (Gillis, 2017). Moreover, Geers highlights the difficulty of controlling an ever-growing quantity (Geers, 2010). He nevertheless emphasizes that these are technical challenges that may be solvable with increasing research in this area. In addition, due to the previously described dual-use factor, states do not have sole control over the means used as weapons, but non-state actors also have ownership and operational rights in this domain. Consequently, for an arms control treaty to be effective, actors from the private sector must be involved and should be committed to such an endeavor (Arimatsu, 2012) as well as the industrial sector, as cyberweapons arms control could have a strong impact in the form of high additional effort, cost, or bureaucratic overhead to implement and perform such controls (Denning, 2001) as well as finding the regulation balance that does not hinder the IT security industry. Beside these, cyberspace is also place of individuals or non-state groups, that have expressed power within this domain, either in the field of cybercrime, political or ideological

hacking, or in the context of state conflicts (Sigholm, 2013), thus providing a further challenge to the ongoing debates.

E: Political will: The political will is crucial for establishing arms control measures (Arimatsu, 2012; Dunn, 2005; Maybaum & Tölle, 2016; Reinhold & Reuter, 2019a). Due to the borderless nature of cyberspace, it is a prerequisite that many states participate in such a regime (Maybaum & Tölle, 2016). However, there are several reasons why states could be reluctant to participate in such an arms control regime. First, there is the risk that authoritarian states (e.g., Russia, China) may want to preserve patriotic hackers as a political tool and continue to have the ability to control politically threatening internet content that would be protected under the freedom of expression in democratic states (Litwak & King, 2015). Moreover, states could be opposed to any treaty that restricts the development of offensive cyberweapons, believing that it would also limit their ability to adequately build up their cyber defense (Czosseck & Podins, 2012; Denning, 2001; Dunn, 2005). Since it is not possible to build a strong defense without knowing what kind of attacks are possible and what vulnerabilities could be exploited, many states would find their cyber defense hampered by cyber arms control that would limit research into attack methods and tools. Furthermore, there are concerns about whether a regime would be wanted by states considering the cost-benefit calculation of cyber arms control since the cost of enforcing and monitoring a global ban may be higher than the expected reduction in risk (Arimatsu, 2012). All this is made even more complex by geopolitical tensions, mistrust, and divergent interests which complicate negotiations on international cooperation (not just) in this area (Hansel & Silomon, 2021).

The overall view of these challenges to arms control in this domain discussed in the scientific literature gives the impression that arms control is failing in face of the realities of cyberspace (Reinhold, 2019c). Consequently, new forms of transparency and verification are needed specifically for cyberspace, as well as qualitative rather than quantitative arms control. It seems that the conventional methods of arms control to ensure transparency and verification have had their day against the backdrop of cyberspace (Altmann, 2019b; Ruhmann, 2015). Therefore, several researchers conclude that cyber arms control will not be possible (e.g. (Maybaum & Tölle, 2016)). At the same time, despite all these challenges, other authors (Altmann, 2019b; Denning, 2001; Dunn, 2005; Litwak & King, 2015; Meyer, 2011) emphasize that arms control in cyberspace is urgently needed. Moreover, international understanding of the danger of uncontrolled militarization of cyberspace is increasing (Reinhold & Reuter, 2019a). General media has also been addressing the issue for several years with, e.g., articles in the New York Times (The Editorial Board, 2015) and the Economist (Economist, 2010) advocating for cyber arms control. In 2014, Meyer argued that cyberspace had not yet become an active battleground for cyberwar at that time, but he stressed the possibility that this could change soon (Meyer, 2011). This can currently be observed regarding the Russian war on Ukraine, where cyber measures play an important part in Russia's hybrid toolkit (Tidy, 2022). In the spirit of Meyer (Meyer, 2011), to face these challenges, now is the time to adopt an arms control approach to cybersecurity as a measure for conflict prevention and mitigation. In doing so, he points to experience showing that preventive strategies regarding new threats have proven to be a more efficient and effective means of combating them than attempts to retroactively contain threats that have already emerged. He points out that to do so, lessons should be learned from the past and the extensive

experience of arms control. According to him, the inventory of previous arms control models is extensive and flexible enough to meet the specific challenges of cyberspace.

Research Gap: As discussed in the previous sections, the new (military) domain – cyberspace – is an unstable environment without explicit agreements among states, which "invites miscalculation, misinterpretation, and inadvertent escalation of conflict" (Lewis, 2013). The increasing risks in cyberspace pose a challenge to civil, political, and military security and stability. Hence, cyber operations have the potential to threaten international peace and security. However, the domain lacks clear and binding agreements due to complex challenges. Therefore, it is necessary to clearly identify these challenges to further analyze which instruments or measurements are most effective to change this. To satisfy this need for clarification, this study does not rely solely on the literature presented above, but aims to take a more holistic approach by incorporating the opinions and experiences of various experts in the field.

9.3 Methodology

In the following, the methodology employed in this study is described in more detail. First, the data collection, the choice of interview partners, and how the interviews were conducted will be explained. Second, the method of the qualitative content analysis is presented, which was chosen to evaluate the collected data.

9.3.1 *Data Collection*

There are policy areas that are discussed little in the broader public, but predominantly in small expert circles – security policy, especially cybersecurity or arms control, is such a policy area (Gais, 2019). This is because much information is sensitive and secret and not intended for the public. Furthermore, security policy topics are very far away from the experience horizon of the general population (Biehl & Jacobs, 2014). Even in parliamentary circles, security policy is considered an expert issue (Rüger, 2012). Thus, the chosen topic of the paper represents an expert topic, i.e., a field that is not easily accessible and for which specific knowledge is required. Therefore, to answer the research question and to generate specific knowledge, experts in this field were relevant points of contact in the context of this article to collect data. The selection of interview partners was based on their expertise relevant to answering the research question (Meuser & Nagel, 2009), evaluated based on their publications or by personal experience with the experts. This includes experts, such as researchers, from different fields of arms control and cybersecurity, and individuals involved in policy processes related to these topics. The inclusion of individuals from these fields was intended to ensure the necessary interdisciplinary lens of the topic and to ensure the broadest possible consideration and inclusion of diverse perspectives. In total, we conducted 10 interviews in December 2021 and January 2022 with experts from Germany, the U.S., and Switzerland, covering a broad range of experts. The interviews were conducted based on a semi-structured guideline that was created based on the qualitative literature analysis of the challenges for arms control in cyberspace presented

in chapter 9.2[1]. The guideline serves to provide a certain framework for the interviews and the data obtained from them (Mayer, 2012). It ensured a clear thematic focus within the interviews as well as the thematization of all important aspects. Further, it allowed for a certain degree of comparability (Kruse, 2015). At the same time, the semi-structured nature of the guideline allowed to ask follow-up or ad hoc questions and to change the order of the questions during the interviews. Although the sample is rather small, with 10 interviews, we are confident that our methodology has allowed us to include a broad and diverse range of different expert views, a perspective that is backed up by research like that of Caine (Caine, 2016). The interviews were individual guided interviews conducted face-to-face online. The audio recordings of the interviews were subsequently transcribed in verbally smoothed form, as only the content of what was said was relevant to answering the research question, not para- or nonverbal expressions. Therefore, the spoken language was converted into written language, following the transcription rules according to Kuckartz (Kuckartz, 2018). To encourage an open and candid conversation, the interviews were anonymized during transcription and evaluated anonymously. Consequently, no data that could be linked to participants, e.g., names or institutions, were disclosed as part of the evaluation.

9.3.2 *Data Evaluation*

To evaluate the knowledge generated through the interviews in a structured way, a qualitative content analysis was conducted, using Mayring (Mayring, 2015) as a basis for orientation. Additionally, Gläser and Laudel (Gläser & Laudel, 2010), who build on Mayring's approaches, were considered in the context of this work to increase openness and flexibility in dealing with the material. The procedure is rule-guided and thus comprehensible as well as verifiable. Further, it is theory-guided, as theoretical preliminary considerations form the basis for the evaluation criteria (Mayring & Fenzl, 2014). To analyze the material, a category system was formed deductively from theory, which was further defined in a coding guide resulting from the qualitative literature analysis so that systematic links were made to the state of research. At the same time, the openness of the qualitative approach was used, which meant that further categories and codes could be inductively added from the material during the coding process to ensure that characteristics that did not fit into the predefined search grid were also considered in the analysis. Thus, the category system was formed on the basis of the interrelationship between theory and data, while maintaining the tension between theory and data. For this purpose, 13 codes were derived from the theory in advance. Moreover, six codes were added during the coding process. In total, all codes were used except for three that were too broad, hence, more specific codes were a better fit[2]. The interview transcripts were evaluated by searching for relevant information using the codes as an analysis grid, assigning this information to the categories, and thus extracting information from the texts in a systematic procedure and presenting the content structure. During the qualitative content analysis, all interview transcripts were examined, whereby the order in which the transcripts were examined was irrelevant to the analysis. The direction of the analysis was determined by the research question. Therefore, the information in the transcripts was of interest, not the person who expressed it. The material was

[1]The interview guideline can be found in the appendix in 9.7.2
[2]The coding scheme can be found in the appendix in 9.7.3

analyzed in a structured manner so that the focus of the analysis was particularly on filtering out certain aspects of the material that were relevant to answering the research question and were identified in advance through the qualitative literature analysis. Single words may have been coded as the smallest text component (= coding unit), as these may be important keywords or key terms, up to the entire answer to a question as the largest text component (= context unit). Multiple coding of individual text passages was possible. The results and interpretations were compiled against the background of the research question (Mayring & Fenzl, 2014). The text analysis software MAXQDA was used to code the interviews. Due to the limited scope of the study, it was not possible to test the analysis regarding the content-analytical quality criteria according to Mayring, such as intercoder reliability, whereby a second person is consulted and codes the material (Mayring, 2015).

9.4 Empirical findings

In the following, the results of the qualitative analysis of the expert interviews on the challenges to establishing arms control in cyberspace are presented. The analysis aims to critically reflect the findings from the literature analysis and to identify core challenges for further discussion.

9.4.1 *A: Lack of a definition for the term cyberweapon*

The lack of suitable definitions emerged as a central challenge during the interviews and was mentioned by more than half of the experts interviewed. One expert emphasized *"I don't think we can actually quantify or define what in fact is a cyberweapon. Is it a computer or a malicious code or what is it? I think that is the first issue"*. With uncertainty on how such a weapon or instrument of attack could be defined in the first place, some experts concluded that this is an unsolvable problem or agreed that it is not solvable at present: *"The main challenge is, that there is no clear definition of what a cyberweapon even is. It is not something that we can specify or define to the point"*. Notably, some of the experts stated to be critical of the term cyberweapon or even went so far as to say that, in their view, a cyberweapon as such does not even exist:

> "First of all, I have a problem with the word 'cyberweapon'. Because a weapon for me is defined in some form a kinetic use of energies. I don't necessarily have that with a cyberweapon, whatever that is then. It's just a script. (...) For me, there are no cyberweapons in that sense. There are exploits, there are vulnerabilities, there is exploitable information technology that can only lead to a kinetic cascade effect in the aggregate, not necessarily."

The question of how to define a cyberweapon is particularly relevant because it also depends on what would consequently be controlled by international arms control treaties and frameworks. This is complicated by the fact that there are numerous ways to carry out a cyberattack, including DDoS attacks, attacks based on zero-day vulnerabilities, or

computer worms. This leads to the fact that cyberweapons are not covered by the classic definition of a weapon:

> "In the case of cyberweapons, the mechanism of action is of course not as direct as in the case of conventional or nuclear weapons, because the cyberweapon does not act against humans, but against machines or against control systems. Nevertheless, we have already had experiences where we can see that the effect of a cyberweapon can also be equated with a conventional weapon. So, whenever it's about sabotaging or destroying critical infrastructure."

One of the experts also pointed out that the discussion about cyberweapons covers a very broad spectrum, which also makes it difficult to find suitable definitions. For example, a cyberweapon can be used for espionage, which would be largely legitimate, or to cripple critical infrastructure, which could in the worst-case scenario cost human lives due to cascading effects. Notably, another expert highlighted that it is not impossible to find suitable definitions. He emphasized that it is *"extremely challenging because we can hardly say on the basis of a technical specification what is good and what is bad."* Consequently, finding suitable definitions is complicated by the dual-use aspect prevalent in cyberspace.

9.4.2 *B: Dual-Use-Dilemma*

The dual-use dilemma that prevails for IT soft- and hardware presents another frequently mentioned challenge. For example, some experts noted that the basis of cyberweapons are codes and software that can be used for different purposes: *"One of the big challenges is that, yes, cyberweapons are not declared as such, but they are software that by definition can be put to different purposes"*. In this context, techniques that are used for an attack can at the same time serve the legitimate purpose of securing national infrastructures and thus form an important tool for maximizing national cybersecurity: *"[in cyberspace], it can be that no matter what it is, a ready-made software or just a piece of code or the knowledge around an exploit, that can become a weapon or just a useful, useful, helpful tool to administer or improve anything"*. The dual-use factor thus not only describes both civilian and military applications, but also includes offensive, defensive, scientific, and industrial application. In particular, the lack of distinction between the militarization of cyberspace, such as the development of cyberweapons, and cyber tools used and developed for espionage purposes poses a major challenge in the eyes of some experts. This was particularly emphasized as both use cases basically rely on the same soft- and hardware tools as well as the same expert knowledge of vulnerabilities and how to exploit them. This aspect therefore also complicates verification in this area. Moreover, these attack tools can be used by a variety of actors, such as hacktivists, making cyberweapons very easy to spread.

9.4.3 *C: Verification*

Finding suitable verification mechanisms to establish arms control in cyberspace is an extremely difficult, but at the same time essential challenge. One of the experts working on this topic in a political institution emphasized: *"This is always held against us, about 'we can't verify this, so it's not going to do any good'"*. For example, cyberweapons cannot be quantified. Accordingly, it is not possible to count weapons or ban an entire category of weapons, as it has been the case with arms control agreements in the past. Nor do they require large industrial facilities, etc., to produce them; a laptop alone can do the job, as one expert stated: *"this kind of simple, countable, measurable, clear, unambiguous verification we probably won't get in these fields."* Additionally, it is possible to infinitely replicate cyberweapons and send them all over the world without cost. Even unintentional proliferation can play a role in this domain, which means that even attackers themselves can never be sure that their capabilities will not be reused or expanded by others: *"With code; just because you delete it of a device, it does not really mean it is gone – most likely it is probably somewhere else, on either some forms of backup system or the internet."* This aspect also exacerbates the challenges of establishing suitable verification mechanisms, as they would have to be extremely intrusive. Thus, challenges concerning privacy confidentiality, proprietary information, and privacy information could arise. Against this background, the challenge arises that states would likely be unwilling to participate in verification mechanisms in this area, since they would also have to provide insights into their cyber defenses for verification purposes. Thus, the danger could be seen that these insights could be misused to spy on vulnerabilities. As highlighted above, the issue of verification is further strongly related to the challenge of missing definitions, according to which it is unclear what should be verified in the first place.

9.4.4 *D: Constant technological development and role of the private sector*

Another challenge that was expressed by the experts on establishing arms control in cyberspace is the ongoing technological progress in cyberspace. Enormous momentum continues, not only in military terms, but also technical developments of cyberspace and its infrastructure in general. It is hard to predict where this development will go, when it will end, or when it will slow down:

> "While the cyber sector, not only in the military sense, but also in the whole IT development, is still characterized by an enormous dynamic and it is not yet foreseeable where the whole thing will go. This makes it much more difficult, I believe, to develop such in-built arms control mechanisms in such a phase, where things are still very much in development, and it is not yet possible to foresee everything that is still to come. I think the great challenge in arms control is always to develop arms control policy steps when the development of this technology or type of weapon is still in full swing and is therefore very dynamic"[3].

[3]Translated citation. Answer was originally given in German.

Additionally, some countries like Russia or China are showing strong interest and initial activities in decoupling from the so-far common technological developments in this domain or to pursue own developments with national led interest like, e.g., the NewIP[4] (Godehardt & Voelsen, 2020). In such a phase, it is extremely difficult to establish arms control mechanisms that have a stabilizing or limiting effect, as states simply have no interest in doing so either to sustain current national advantages or to gain an advantage. Related to this is the fact that the code of a cyberweapon is usually based on ongoing software developments that are extended and adapted for a specific target and task and therefore evolve very easily and quickly. Accordingly, the possibility of variation is extremely high, and that future cyberweapons will always be (somewhat) different from past cyberweapons. This complicates any kind of regulation for arms control and verification measures that are based on technical features of a malicious software tool. Due to the discussed dual-use factor, as well as the aspect that most of the relevant cyberspace infrastructure is privately owned, the private sector needs to play a relevant role in establishing arms control, especially for the implementation of verification measures for controlling and enforcing agreements that – even if not yet developed – usually need some kind of technical measures or adjustments to existing systems (Reinhold & Reuter, 2019c). Moreover, the private sector is the primary provider for most data and information like the knowledge of vulnerabilities but also threat information and threat hunting know-how:

> "Basically, the infrastructure is mostly privately owned. That means that even when we talk about the distribution of cyberweapons, e.g. malware, it is not released through state means, but through private infrastructures. And we have to work together with the private actors, the owners of the infrastructure, the telecommunication providers, also with the cybersecurity companies, who have to find their role in this, in order to control the use of such software. All these would also have to be involved in the verification mechanism."[5]

Finally, the private sector would be strongly affected by regulation measures and their requirements need to be considered as not to jeopardize further development in the area of IT security.

9.4.5 *E: Political will*

Another challenge that the experts considered to be central is political will. The experts pointed to the close connection between this and the dynamics already described, which are currently shaping cyberspace. Although there have been some small steps towards a common understanding, especially regarding the validity of international law norms in cyberspace that have been confirmed by the UN Group of Governmental Experts on Advancing responsible State behavior in cyberspace in the context of international security - UN GGE (Schulze & Datzer, 2021) and the UNODA[6] Open ended working group on security of and in the use of information and communications technologies

[4] The initiative is sometimes also called IPv6+.
[5] Translated citation. Answer was originally given in German.
[6] United Nations Office for Disarmament Affairs

- OEWG (UNODA, 2022), according to some experts, states are currently still in the discovery phase as to what advantages and opportunities cyberspace could offer to them. As a result, states are just beginning to perceive cyber tools as strategically valuable and have diverging interest, even between states that otherwise share common values and interest, like in the EU (Wisotzki & Mutschler, 2021):

> "Countries are still in the exploration mode, where they're trying to understand what the boundaries are, what is doable, where is their offense advantage, where the defensive. Governments are not going to want to give up because either they are becoming really useful to them in the future or they're afraid that their competitors will cheat and lie and even if they say they won't sign the treaty. Or if they sign the treaty they'll still cheat and they'll get some huge advantage from having a cyberweapons capability that you won't have if you're complying with the treaty and therefore, you'll be at a big disadvantage."

The many possible use cases of cyber instruments already described are considered highly relevant and worthwhile for states, as they do not want to forego the associated advantages, and especially espionage activities tend to become unexpected norm setters for state behavior in this domain (Georgieva, 2020). This can be seen, for example, in the fact that cyber tools are already being widely used by various states for these purposes and that companies that buy and trade the knowledge of vulnerabilities in IT hardware and software are growing, with a focusing on state actors and agencies as their primary customers. Furthermore, states would probably not consent to any agreement that does not include either functioning or overly intrusive verification mechanisms – the mistrust that other states would not comply with such a ban is currently too great, and the fear of accepting disadvantages vis-à-vis competitors prevails. Furthermore, the overall geo-political situation makes progress seem a distant prospect. Finally, a last challenge in this context is that cyberweapons can be used by a variety of actors. As described earlier, they are not exclusively in the hands of a state. Accordingly, a variety of actors are relevant for the restriction of the use of cyberweapons but probably not addressable as arms control agreements are concluded exclusively between states.

9.4.6 *Further remarks from the interviews*

An important result of the analysis is that political challenges, such as the lack of interest on the part of states to agree to such a convention, and the lack of relevant definitions were coded much more frequently than other challenges. Thus, these two challenges can be identified as particularly central according to the experts interviewed. In addition, the dual-use dilemma and the difficulty of finding suitable verification mechanisms were highlighted in the interviews. At the same time, however, it is interesting to note that the experts weighted certain challenges differently. For example, one expert named the lack of suitable definitions and verification mechanisms as the most central challenge, while another focused particularly on political will as a challenge. It is also noteworthy that the role of the private sector was primarily discussed by experts working in a political institution. These differences show that it is important to speak with experts

from different relevant fields about this topic, as this allows different perspectives to be incorporated into the analysis.

9.5 Discussion

9.5.1 *Discussion of the Results*

In general, the main challenges identified through the literature review in the theoretical part (in chapter 9.2), that are also briefly shown in Figure 9.2 in the appendix (section 9.7.1), were also discussed by the experts interviewed. However, a differentiation can be observed in the weighting of the individual challenges. It is noticeable that the lack of political will is seen by the experts interviewed as a more central challenge than is the case in the academic literature. In some cases, statements even contradict each other: Geers, for example, analyzed in 2010 whether the mechanisms of the Chemical Weapons Convention could be applied or transferred to cyberspace (Geers, 2010). As a result of this analysis, he emphasized that he saw the cyber threat and the danger that terrorists could use this sphere to achieve their goals as an opportunity strong enough to build political consensus. At the same time, it is critical to note that not only more than a decade has passed since Geer's analysis, but much of the academic literature used in this paper was published between 2001 and 2016. However, these are works that have been cited more frequently and are more relevant in this sense than others. Based on this, two interesting aspects can be observed: On the one hand, this suggests that the political climate with regard to international cooperation has deteriorated. This coincides with the frequent description of geopolitical tensions and a crisis of multilateralism in recent years, which complicate international cooperation (Brühl, 2019; Munich Security Conference, 2019; Neuneck, 2018), as well as the crisis of arms control often mentioned in academic literature (Becker et al., 2008; Daase et al., 2019; Meier, 2020; Nassauer, 2008).

At the same time, a key finding of the analysis is that the experts interviewed differentiated between the individual challenges in similar contexts: For example, challenges such as the lack of suitable definitions or political will often were mentioned when the experts were asked about the main challenges for cyber arms control. Only when it came to questions regarding an actual implementation of arms control measures did further challenges come up in connection with the difficulties of establishing suitable verification mechanisms or difficulties, which were discussed against the background of the dual-use factor. Thus, the challenges can be divided into structural challenges, which are caused by the structure of cyberspace, and procedural challenges, which become relevant in the actual process of establishing cyber arms control (see Figure 9.1).

The procedural challenges are particularly relevant to answer whether cyberspace can benefit from mechanisms established within the domain of chemical weapons. In addition, these procedural challenges also raise the question whether they could also be solved by technical solutions or if certain problems lie beyond technical possibilities. Another related finding of the analysis is that the experts' assessments of whether and which challenges can be overcome differ greatly in some cases. It should also be noted that challenges were often not explicitly named but were roughly paraphrased or de-

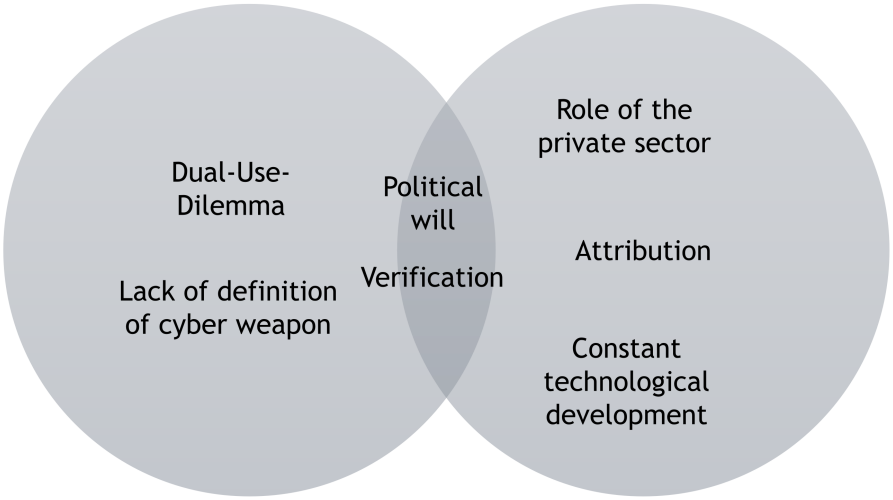

Figure 9.1: Classification of challenges into structural (left) and process-related (right) challenges. Source: Own illustration.

scribed by the experts. In addition, some of the challenges became somewhat blurred in the conversation. For example, the challenge of the lack of definitions was not always explicitly mentioned, but was described during questions regarding the dual-use nature of cyberspace. This illustrates how closely interrelated the individual challenges often are: The issue of verification, for example, is strongly related to the challenge of missing definitions, since without specific and consistent definitions it is unclear what should be verified at all. Also, the discussion of the results shows that several challenging aspects were often subsumed under a given term, which makes finding solutions even more difficult. At the same time, however, this also shows that individual aspects can be discussed against the background of various different overarching categories: For example, the dual-use factor of cyber instruments is in itself a major challenge for establishing arms control mechanisms in cyberspace. At the same time, it also makes it difficult to find suitable definitions or appropriate verification mechanisms.

9.5.2 *Interpretation*

The analysis leads to the conclusion that, according to the literature and the experts, neither the control of a cyberweapon nor any other technological regulation for cyberspace will work. Instead, the focus must be on banning certain actions, since the experts do not see any chance for verification mechanisms, especially because of the high level of intrusion that would be required. This is in line with Roßner's concern (Roßner, 2017) that these very control measures, which are supposed to compensate for a lack of trust, may in turn trigger mistrust, as actors may fear that they are being spied on by weapons inspectors. The analysis suggests that traditional measures of arms control cannot be transferred from one area to the area of cyberweapons. Instead, it is necessary to create new alternative and creative solutions for the domain. Considering this, the analysis

shows that one possible solution could be to define and sanction not the weapon itself, but rather certain uses of the tools that could be prohibited, an approach that has, e.g., also been expressed by Hansel et al. (Hansel et al., 2018) based on the experiences with preventive arms control methods. Such an approach could help to overcome the challenges highlighted by the lack of definitions, the dual-use dilemma, as well as continued technological development. Thus, the criticism expressed in the literature by Roßner (Roßner, 2017) that existing treaties are often tied to technical characteristics of specific types of weapons, which makes them blind to new types of weapons due to low levels of abstraction would be overcome. This would also allow agreements to be made independently of the pace of development of the area, would be technically more feasible and still represent a security gain. Such a behavioral regulatory approach has already been used to create norms for responsible behavior in cyberspace (Schulze & Datzer, 2021). Nevertheless, considering what is politically feasible, the analysis shows that the challenge of a lack of political will still pose a major problem for implementing binding rules. This result can therefore be discussed against the background that for years many experts have been writing about a crisis in arms control (Becker et al., 2008; Daase et al., 2019; Meier, 2020; Nassauer, 2008) triggered by the crisis of multilateralism (Brühl, 2019; Munich Security Conference, 2019; Neuneck, 2018). Hence, an expert expressed the idea that right now, he only sees soft law options in the normative realm as a possibility. Moreover, it must be considered that a vacuum in international law could arise in the period between the negotiation of a cyber agreement and its entry into force, which means that this phase of negotiation, the goal of which is to increase security, could give rise to a phase of uncertainty. Another objection is that a potential arms control negotiation would, like any arms control agreement, result in an intergovernmental treaty between states regarding their military capabilities (Reinhold, 2019c). Such a treaty-based approach has its limitations, especially regarding the mentioned role of non-state actors that can – at best – only be addressed indirectly via national legislation, law enforcement and by fostering and strengthening the due diligence principle of state responsibility. Nevertheless, this paper concludes that an intergovernmental agreement regarding cyberspace would still represent a security gain, especially since states are the main source of danger in cyberspace because they primarily have the capabilities as well as resources for large-scale cyberattacks (Lewis, 2013). Such binding rules could also lead to fewer states accepting non-state groups carrying out such attacks on their territory and thus no longer being considered safe havens. The analysis confirms that the experience of a threat has a strong influence on the will to negotiate. The experts assumed that experiences such as devastating cyberattacks would change the willingness of relevant actors to reach agreements. The motivation for the establishment of an agreement would therefore be a changed perception of danger. This is in line with Roßner (Roßner, 2017) arguing that a major shortcoming of arms control agreements is that such treaties are often discussed only after major damage has already been done by the relevant weapons, and thus agreements of this type often come too late. Likewise, this result confirms the assessment of Maybaum and Tölle (Maybaum & Tölle, 2016) that a common policy is possible when civil societies recognize an imminent threat beyond national borders. Even though the strategic utility of cyberweapons is not to be underestimated (CCDCOE, 2022), the cost-benefit considerations of regulating cyberweapons are currently to the detriment of an agreement. This makes it difficult to develop political will for an agreement. However, the analysis also shows that the perception of the strategic value of certain instruments can change over time. An important finding is also that it is entirely possible to overcome technical challenges over time through research. This is in line with Geers (Geers, 2010) emphasizing that

technical challenges may be solvable with increasing research in this area. For example, attribution – the forensic and political process of collecting secure knowledge about the origin of a cyberattack – was considered a much bigger, if not unsolvable, problem about 10 years ago, but is now performed regularly and thus represents an important tool for arms control to determine the use of a certain cyberweapon and its origin. Regarding further technical challenges of a cyber arms verification, it cannot be ruled out that solutions for this might be found in the future. In 2010, Lewis emphasized that we were still in the very early stages of thinking about how to create cybersecurity as a global community. Moreover, Meyer wrote in 2014 that it may be too early to establish arms control treaties for cyber instruments, given the challenges that currently remain. The analysis leads to the conclusion that we are still in this early stage:

> "The big challenge in arms control is always to develop arms control policy steps when the development in the field of this technology or type of weapon, genre of weapon is still in full swing and therefore with great dynamics." Nevertheless, cybersecurity is gaining relevance, albeit slowly: "As far as cyber is concerned, I would think that the concern of states and societies, their vulnerability, recognizing that, that's just happening now. That's basically a consequence of the digital transformation. What other consequences this has, including security policy consequences, when a society is digitally transformed is only now slowly becoming clear."

Overall, the analysis shows that little has changed in terms of arms control challenges in cyberspace since Meyer's observation in 2014. Nevertheless, the idea or goal of cyber arms control should not be dismissed prematurely. Therefore, more research is needed because arms control still represents a successful project of the past and thus an important instrument of international relations to create more security.

9.5.3 *Limitations and Future Work*

Finally, some limitations regarding this work should be mentioned. Although the selection of experts was based on theory-based research, it was also dependent on the availability and willingness to be interviewed. Also, most interviewees had a background in political science, and only few in science or technology. This is relevant given the interdisciplinarity required to address questions in this domain. This lack of technical expertise possibly leads to misperceptions of technical limitations of cyberspace and the reproduction of techno-pessimistic perspectives. It should also be critically noted that more men than women and exclusively persons from the Global North were interviewed. Moreover, individual opinions were collected, which is why generalization is not possible and the results neither represent an expert consensus nor did we aim for comparability and prioritization of challenges, for which a questionnaire would have been necessary in contrast to our open-ended question approach. Additionally, due to the small circle of experts working on arms control in cyberspace, there was to some extent overlap between the literature we used to illustrate the relevance of the research topic and the expert interviews[7]. In this context, there is also the threat of

[7] In detail this affects the literature that we used from J. Altmann, M. Hansel, K. Geers.

circular reasoning, as the researchers we interviewed are presumably also aware of the limited research available. Nevertheless, as the results show a differentiated range of highlighted challenges and problems, we believe that our chosen approach help to mitigate this threat. In addition, it should be noted that the interviews were conducted before the outbreak of the Russian war of aggression against Ukraine. It is unclear whether these events have changed certain perceptions. This article contributes to the question whether lessons can be learned from combining academic and practical perspectives on establishing arms control. Much more research is needed to comprehensively discuss which paths are useful and viable for enhancing cybersecurity and which are not. To do so, even more diverse perspectives on the topic should be included. More attention should be paid to both the technical perspective and the potential impact of cyberattacks. In addition, various relevant stakeholders, including policymakers as well as the private sector, should have their say and be able to contribute perspectives and expertise in a multi-stakeholder approach.

9.6 Conclusion

Cyberspace represents the fifth space of warfare and is becoming increasingly relevant in conflicts. Thus, cyber capabilities are coming into focus in security policy thinking, e.g., through a corresponding emphasis in state military strategies. This raises the question of how to foster cybersecurity globally. This article examined the challenges for arms control in cyberspace from a theoretical perspective and, in a further step, critically reflected on them by drawing on the expertise of various experts in this domain. The analysis revealed the following: Cyber arms control is confronted with a multitude of challenges. The challenges described in the research literature coincide with those described by the experts interviewed. The main difficulties are the lack of political will, of definitions, and of verification capabilities, as well as the dual-use nature of cyberspace. Other challenges include the multitude of stakeholders involved beyond states. The analysis suggests that a broad definitional approach is advisable for cyberspace. It makes sense to regulate behaviors and outcomes rather than the technology itself and to use such an approach to define the weapon or the prohibited use thereof. The analysis also shows that a reliable attribution mechanism, which involves technical as well as political dimensions, is necessary for such an agreement. It must be certain that violations are detected to publicly communicate the misconduct, trigger possible non-compliance consequences of the regulatory regime, or otherwise deal with it either politically, economically, or even militarily (Broeders et al., 2020). Such a mechanism would need to be independent and credible. However, this is precisely where an opportunity arises, as the analysis shows that cyberattacks have often been reliably attributed in the recent past. Technical challenges have thus already been successfully overcome. It is now urgent to translate the processes initiated under the UN GGE and the OEWG into binding regulations – as unrealistic as this may sound in view of the current world situation.

9.7 Annex

9.7.1 *Overview of the challenges to establishing an arms control agreement in cyberspace*

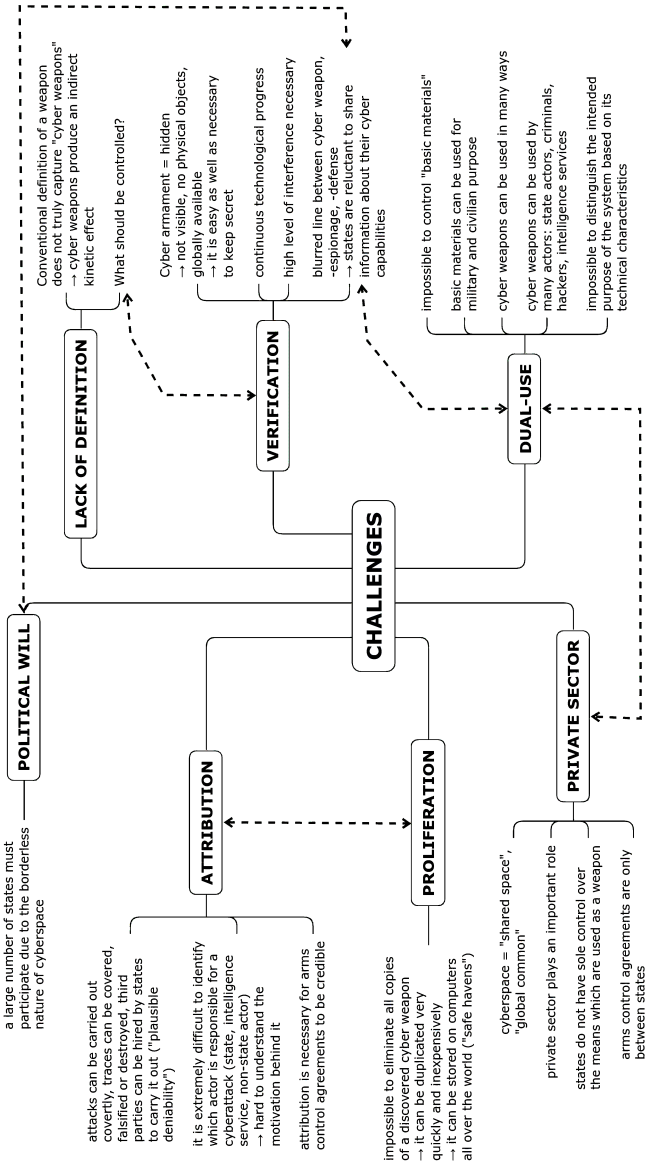

Figure 9.2: Overview of the challenges to establishing an arms control agreement in cyberspace. The dashed lines represent particularly strong connections between the respective aspects. Source: Own illustration.

9.7.2 *Interview guidelines*

Remarks: We used the conducted interviews for two different research projects. The second research project analyzes similarities and differences between arms control for chemical weapons and cyberweapons. The interview guideline scheme therefore sometimes refers to chemical weapons. The guidelines below are a shortened version, reduced to the aspects concerning this research paper.

The role of MHC in autonomous military systems against the background of an actor-systematizing view. In our work, we deal with the concept of Meaningful Human Control (MHC) in autonomous military systems. Here, I am mainly interested in how different actors define MHC for themselves and what implications this may have for a possible implementation. The starting point is that in the debate about MHC there are different ideas about what human control means or should mean – this is not least an obstacle to agreeing on international arms control rules. Nevertheless, numerous states are already developing weapon systems with autonomous or semi-autonomous functions – both within the EU and within NATO, systems are currently being developed that pose a challenge to human control – a prominent example here would be FCAS. With this in mind, I am initially interested in:

Introductory question: What does MHC imply for you and what role do you think does it play in the context of autonomous military systems? Would you say MHC and the autonomization of autonomous systems or autonomous weapon systems represents a contradiction?

The development of autonomous military systems involves a whole range of different actors from civil society, research and development, and business and industry.

Question 1: How would you describe the current (1) technical debate on the responsible and ethical use of AI-driven autonomous systems? What are core aspects that you would highlight here? How does it differ from the (2) policy debate?

Demand I: How would you see the interaction of different stakeholders, for example from civil society, academia, military, or policy, in the implementation of MHC in autonomous systems development to date? Whose role in Research and Development do you rank as the most influential and why?

Demand II: How useful would you consider a multi-stakeholder approach to the development of autonomous systems? Which stakeholders should definitely be brought in?

In our research, we examine the websites of arms-producing corporations for narratives on MHC in autonomous systems. Thereby the impression is created that the defense industry promotes the use of autonomous systems in a military context with many different narratives – one example is that operating in a "contested battlefield" or "contested environment" makes it indispensable for military to deploy a "human-machine team". Autonomous systems are thus promoted as necessary in order to be able to act efficiently and quickly and are thus seen as the future of warfare. With these narratives, one quickly gets the impression that the industry side is quickly falling into technical discussions with little consideration of political and ethical aspects.

Question 2: I would be interested in your (critical) assessment of this. Do you see an implementation of MHC in the development of autonomous weapons by the defense industry?

Inquiry: What do you think would have to change in the status quo to establish a broader debate on MHC in the defense industry as well? Do you have any examples?

Question: At the same time, however, companies also emphasize the need for a so-called human-machine team, in which the AI or AI-controlled autonomous system acts primarily as a support for humans. In this context, I brought a statement/slogan that crossed my way while analyzing the website of a defense company that develops autonomous systems. It says: "The Future of Autonomy Isn't Human-Less. It's Human More." How would you evaluate this statement in the MHC context (critically)?

Question 3: What are technical and operational parameters that autonomous systems must meet for responsible and ethical use?

(**Demand II**: How can technical design of autonomous military systems suggest and facilitate accountability? What might this look like specifically in future systems?)

(**Demand III**: Can you give concrete examples where it is possible to speak of a successful implementation of MHC?)

Question 4: Keyword human-in-the-loop, human-on-the-loop and human-out-of-the-loop: What role will autonomous military systems play on the battlefield of the future? What advantages or risks do you see in this? And how many humans will still be involved in warfare in the future?

Question 5: That are all the questions I have. In your opinion, are there any important aspects that we have not yet addressed?

Question 6: Who else do you think we should ask about this?

Why this research question at all? - On the one hand: In various papers or lectures, one hears again and again about (possible) parallels between the chemical weapons domain and the cyberweapons domain. - Parallels can be observed especially in the challenges that these types of weapons pose to arms control: dual-use and proliferation play a major role in both domains, the weapon itself is difficult to define, constant technological progress is expected in the domain and, above all, the type(s) of weapon puts previous verification mechanisms of "classical" arms control to the test or shows that they are not applicable to the domain. - On the other hand: chemical weapons agreement is a great success of arms control, with (almost) worldwide agreement, independent verification regime, surviving crises for arms control (like former US President Trump). Therefore, I came to the idea that it would be useful to analyze to what extent similarities and parallels exist between the two domains and whether and if so, to what extent lessons can be drawn from the past of the Chemical Weapons Convention for a possible cyber arms control.

Question 7: In your opinion, what are the main challenges for a cyber arms control agreement? If cyber arms control is generally not possible, why do you think so? What are in your eyes the unsolvable problems, so that cyber arms control is impossible?

In the chemical weapons domain, similar challenges (or challenges that at least appear similar) to arms control agreements could be observed as currently in the cyber domain. However, new ways of verification have been found to overcome these challenges, in the form of (1) lists of banned chemicals, (2) inspections by an independent specially created verification regime, and (3) through the General Purpose Criterion, which focuses the agreement or definition of a chemical weapon on the purpose or intent of its use.

Question 8: To what extent could such measures of verification be effective for a cyber arms control agreement? To what extent would there be a need for adaptation here? At the heart of the Chemical Weapons Convention is the "General Purpose Criterion" – would such a focus on the purpose or intent of the use of certain means be conceivable or useful for a possible cyber arms control agreement? (Given the difficulty of defining a cyberweapon and its dual-use nature, but especially given that the Chemical Weapons Convention does not explicitly define the purposes of chemical weapons, but rather specifies many more purposes that are not prohibited by the Convention).

Question 9: By whom could a cyber arms control treaty be controlled? Would creating an independent verification regime be one approach to meeting the challenge so that states would share information related to their cyber capabilities? What functions would such a body need to handle?

Question 10: What do you see as (1) technical as well as (2) political challenges to transferring Chemical Weapons Convention measures to cyberweapons? Possible question (if more generally desired): In your estimation, what are (1) technical as well as (2) political challenges to implementing arms control measures for cyberweapons?

Bonus question: In your view, what is the factor or common denominator among states that has made and continues to make a common policy on chemical weapons possible (or arms control in the past), or what is missing or needs to change so that this is also possible regarding cyberweapons?

Final question: That would have been all the questions on my part now. In your opinion, are there still important aspects that we have not yet addressed? Who else do you think we should ask about this?

9.7.3 *Coding scheme*

Remarks: We used the performed interviews for two different research projects. The second research project analyzes similarities and differences between arms control for chemical weapons and cyberweapons. The coding scheme therefore sometimes refers to chemical weapons. The coding scheme below is a shortened version, reduced to the aspects concerning this research paper.

Code	**Description**	**to be coded when ...**
ARMS CONTROL CHALLENGES	All text passages that thematize challenges for arms control	/
Arms control challenges – cyber domain	All text passages that thematize specific arms control challenges related to the cyber domain	/
Arms control challenges – chemical weapons domain	All text passages that thematize specific arms control challenges related to the chemical weapons domain	/
Attribution – cyber domain	All text passages that thematize attribution as a challenge for arms control in the cyber domain or which elaborate to what extent it represents a challenge	*"attribution has been very difficult in the cyber side for lots of reasons. So, the attribution question is important"*
Attribution – chemical weapons domain	All text passages that thematize attribution as a challenge for arms control in the chemical weapons domain or which elaborate to what extent it represented a challenge	*"Covert, the late noticing, the attribution, how can you even prove that, who was the aggressor. That's not trivial even with chemical weapons."*
Verification – cyber domain	All text passages that thematize verification as a challenge for arms control in the cyber domain or which elaborate to what extent it represents a challenge	*"These are not tangible, physical, material systems that can be counted or weighed and somehow have to be synthesized or bolted together first."*

Code	Description	to be coded when ...
Dual-Use – cyber domain	All text passages that thematize the dual-use nature of cyberweapons as a challenge for arms control in the cyber domain or which elaborate to what extent it represents a challenge	*"Dual-use nature; that is, if we look at, for example, dealing with security vulnerabilities, that, so to speak, the techniques you need to secure yourself defensively are sort of the same techniques as the ones you need to attack."*
Political challenges – cyber domain	All text passages that thematize different political aspects or challenges as a general challenge for arms control in the cyber domain or which elaborate to what extent it represents a challenge	*"I think given the technological developments here, the lots of different ways of cyberweapons can be used militarily, and for intelligence purposes, and corporate espionage purposes means that governments are not going to give up the advantage that they can get out of this and use them as well."*
Definition of the weapon – cyber domain	All text passages that thematize aspects of the general definition of a weapon or challenges in defining a cyberweapon as a challenge for arms control in the cyber domain or which elaborate to what extent it represents a challenge	*"main challenge is, that there is no clear definition of what a cyber arm even is. It is not something that we can specify or define to the point"*
Proliferation – cyber domain	All text passages that thematize the proliferation of cyberweapons as a challenge for arms control in the cyber domain or which elaborate to what extent it represents a challenge	*"Proliferation, even unintentional proliferation, so to speak, even the attackers in many cases can't be sure that their own capabilities won't be further used and built by others - that's certainly much more extreme here than in any other field where we've considered arms control so far."*

Code	Description	to be coded when ...
Constant technological progress – cyber domain	All text passages that thematize the expected constant technological progress in the cyber domain as a challenge for arms control in the cyber domain or which elaborate to what extent it represents a challenge	*"cyber is such a quickly changing domain."*
Role of the private sector – cyber domain	All text passages that thematize the role of the private sector as a player in the cyber domain as a challenge for arms control in the cyber domain or which elaborate to what extent it represents a challenge	*"we just have to work with the private actors, the owners of the infrastructure, the telecom providers, also with the cybersecurity companies who have to find their role in this, in terms of arms control, limiting the use of such software."*
Role of non-state actors – cyber domain	All text passages that focus on the role of non-state actors	*"while cyber instruments can ultimately be used by any hacktivist. As a result, there is then also a completely different spread. This is a problem that we are currently facing in the Ukraine conflict."*
Further challenges for arms control – cyber domain	All text passages that thematize further challenges to arms control in the cyber domain	*"What you always have in principle with disarmament agreements is the challenge of also getting on board those states that don't feel that they are affected by it at all"*
VERIFICATION MEASURES	All text passages that thematize specific or possible verification measures	*"the principle of managed access in verification, particularly in suspect inspections, was newly developed at that time to make such a tool possible in the first place."*
Further options	All text passages that focus on further options or alternatives to establish agreements for cyberspace	*"What's realistic would be an agreement between all the countries that have such capabilities and make a clear agreement on what targets are acceptable and what targets are not."*

Code	Description	to be coded when ...
Lists for differentiating prohibited and permitted components	All text passages that thematize lists for differentiating prohibited and permitted components as a specific or possible verification measures	*"This declaration is related to list; that is, there are three lists where certain chemicals are on it. It's based on how dangerous they are from a chemical weapons perspective."*
Verification regime	All text passages that thematize a verification regime as a specific or possible verification measures	*"In the cyber domain, verification would require that you actually have a globally recognized attribution mechanism."*
Verification regime – challenges	All text passages that thematize challenges for the establishment of a verification regime	*"It's more a question of political will today whether we want to create such a place."*
Verification regime – requirements	All text passages that thematize requirements a verification regime would have to meet	*"Further, this body would need to be endowed with credibility – the technical as well as the political credibility"*
Focus on intent	All text passages that thematize a focus on the intent of a weapon as a specific or possible verification measures, to be understood against the background of the general purpose criterion	*"First of all, with the chemical weapon, that approach is not to sanction the weapon or the instrument itself, but to sanction the behavior in using that instrument, that's also the approach that needs to be taken with cyberweapons, if you want to call them that. So behavioral regulation, not technology regulation."*
Focus on intent – challenges	All text passages that focus on challenges against this backdrop	*"I think, you can't have the General Purpose criterion without being able to find out who the perpetrators are and who the persons are behind it."*

Code	Description	to be coded when ...
PARALLELS – CHEMICAL WEAPONS DOMAIN AND CYBER DOMAIN	All text passages that thematize parallels or similarities between the chemical weapons domain and the cyberweapons domain	*"The bigger issue here is to narrow down which aspects of the parallels can actually be investigated"*
Transferability	All text passages that thematize the transferability of chemical weapons arms control measures to cyber arms control	*"the general purpose criterion. Because I don't see how you could implement prohibitions in the cyber area in any other way. So from that, I think the purpose or intent - difficult to determine, of course - could be a useful thing."*
Adaptation requirements	Text passages that thematize specific requirements to adapt measures of chemical weapons arms control to cyber arms control	*"But that's probably where you'd also have to try to have a lot of transparency, a lot of information."*
Similarity	All text passages that thematize similarities of the chemical domain to the cyber domain	*"Chemical weapons, after all, are good for surprises, and that's the same in the cyber domain. Just the covert, the late noticing, the attribution, how can you even prove who the aggressor was."*
Difference	All text passages that thematize differences of the chemical domain to the cyber domain	*"for most military chemical weapons programs you need to have a large infrastructure. You need to produce hundreds if not thousands of tons of chemical warfare agents, you need to fill them into munitions. For cyber, it could literally be one laptop"*
CYBER ARMS CONTROL – RELEVANCE	All paragraphs that focus on or discuss the relevance of arms control in cyberspace	*"a scenario where you have a cyberattack with massive effects on the civilian population, it would of course be terrible, but it could also be a formative event that, from that point on, an awareness arises not only in the expert community, but in general that this is morally unacceptable."*

Table 9.1: Coding scheme

PUBLICATIONS
PART B

10

TOWARDS A CYBERWEAPONS ASSESSMENT MODEL – ASSESSMENT OF THE TECHNICAL FEATURES OF MALICIOUS SOFTWARE

ABSTRACT The revelation of the Stuxnet malware in 2010 shed light on the presence of state actors that are willing and capable of developing and using highly sophisticated, specialized malicious software for their political interests. These tools – often dubbed cyberweapons – are expected to become the next major advancement in weaponry technology. Besides the threats of offensive cyber operations for civil IT systems due to the interconnected nature of the cyberspace, international regulation of cyberweapons is – among other aspects – hindered by the fact that the military development and the strategic and tactical deployment of cyber weapons differ significantly from other weapons technologies. In order to establish measures of cyber arms related control treaties, it is crucial to identify these particular characteristics. Based on this premise, the article analyzes the current perspectives on cyberweapons, identifying their weaknesses of being either based on assumptions about adversarial actors or being applicable only after the usage of a malicious tool. In contrast to these approaches, the article focuses on the specific functional aspects of malware and presents an indicator-based assessment model based on parameters that can be measured prior to the application of malicious software. This enables the categorization of malicious tools as cyberweapons. Besides this, the article aims to introduce thought-provoking impulses with regard to social responsibility in computer science.

ORIGINAL PUBLICATION *Reinhold, T., & Reuter, C. (2021). Towards a Cyber Weapons Assessment Model – Assessment of the Technical Features of Malicious Software.* IEEE Transactions on Technology and Society*, 3(3), 226–239. https://doi.org/10.1109/TTS.2021.3131817*

10.1 INTRODUCTION AND RESEARCH QUESTION

Over the last years, an increasing number of states have included cyberspace into their national security strategies (UNIDIR, 2013) and their military planning (Zeadally & Flowers, 2014). A central element within these developments are "cyberweapons", the technical tools that can be used in the cyberspace for operations against foreign IT systems. Even the use of this term is controversial, because it has legal implications, especially in international humanitarian law, and so far, no internationally unified and binding definition exists. The concerns about an appropriate perspective on cyberweapons could easily be mistaken for a merely theoretical debate. Malware has been extensively researched and many important proposals for its classification and categorization have been made (Ding et al., 2019). Nevertheless, an applicable and

efficient definition of cyberweapons as a subset within the broad range of malware plays an essential role in the regulation of these destructive tools in international relations and the peaceful development of the cyberspace (Robertson et al., 2020). This is especially important for arms control measures such as export regulation, the prevention of unhindered proliferation (Brockmann, 2019) or treaties defining the dos and don'ts of the military application of such tools. Moreover, common criteria for these tools can further help to foster multilateral threat intelligence sharing platforms (Wagner et al., 2019). As this article shows, current approaches concerning definitions or classifications of cyberweapons are mostly either application- or actor-centric and concentrate on the intention or the deployment of malicious IT tools. These approaches perform sufficiently when applied after a specific incident but fail in situations where it is necessary to decide about the weapon character of a cyber tool prior to its usage. This is, at its core, the essential challenge of an effective restriction and monitoring of specific military cyber technologies (Reinhold & Reuter, 2019a).

10.1.1 *Research Question and Methodology*

This paper seeks to examine the following research question: **How can cyberweapons be differentiated within the complex and diverse landscape of malicious software based on features that are determinable without an assessment of their application context or any previous usage?** Our approach follows established arms control measures and looks for the critical components and thresholds that transform a technology or a specific item into a weapon. Such assessments have been established for non-cyber technologies over the last decades. This applies especially in the context of export controls, where manufacturers have to provide technical details on critical – potentially military – goods in order to get an export permission by assigned authorities. The authorities in turn analyze and compare these goods against thresholds and laws that have been defined and negotiated by international treaties and put into national legal norms. Our methodological approach is based on the finding, that after malicious cyber incidents, analytical assessments of the activities and especially the detected malware samples are carried out and published, that focus on the technical details such as code properties and capabilities, exploited vulnerabilities, applied third-party libraries, similarities to existing malware samples, etc. Such technical details on cyber tools are – at least potentially – also available before their use and could therefore be assessed. Following this premise, our approach identifies the technical properties of the development and deployment of malicious software that can be measured "in situ". The findings are compiled into an assessment model for cyberweapons based on a set of analytical indicators as a foundation for arms control measures for the cyberspace. As a first step, Section 10.2.5 presents the range of existing approaches for the classifications of cyberweapons and highlights the research gap. Section 10.3 then analyzes and discusses the technical features of weaponized malware, identifying a set of measurable parameters and indicators. Building on this, Section 10.4 provides our contribution, a suitable and practicable assessment model for cyberweapons, an explanation of how this model can be applied as well as an evaluation based on selected exemplary case studies. Section 10.5 concludes our approach by discussing the assessment model, its applicability and limitations for arms control measures, and presenting further research questions. The Annex will present selected technical details of the case study assessment.

10.2 Related Work: Current Approaches for the Classification of Cyber Weapons

The following section provides an overview of selected works covering the current scientific approaches towards cyberweapons, grouped by their central classification.

10.2.1 *Intent- and Effect-Based Classification*

One of the initial approaches, which is still influential for current debates, has been provided by Rid and McBurney (Rid & McBurney, 2012). The authors dispense with any consideration of specific technical aspects of malicious code but instead focus only on the intention of the application as well as the deliberate selection of the targets by an attacker to distinguish malware from cyberweapons. Their concept already mentions that, besides their intention, malicious cyber tools can trigger unintended or even unforeseen consequences. This aspect of the triggered effects has been further elaborated by an approach of Brown and Tullos (G. D. Brown & Tullos, 2012). The authors propose a spectrum of the actual impact that ranges from non-invasive access and enabling operations, over non-destructive disruptions that suppress a service to destructive attacks. The authors consider the latter as cyberattacks and only the utilized software tools as cyberweapons. The most important approaches in the context of the categorization of cyberweapons were given by the two editions of the so-called Tallinn Manual (CCDCOE, 2013, 2017). Cyber weapons are defined as tools within the cyberspace which are "*means of warfare that are by design, use, or intended use capable of causing either (i) injury to, or death of, persons; or (ii) damage to, or destruction of objects, that is, causing the consequences required for qualification of a cyber operation as an attack*". Although the Tallinn Manual was created by independent researchers for the NATO Cooperative Cyber Defense Center of Excellence, it has become a quasi-official point of reference in international cyberspace politics and its perspective influenced many national security doctrines.

10.2.2 *Classifications Based on the Strategic Assessments*

Some researchers focused on the intent-based approach and the strategic perspective of the attacker and the circumstances of the attack. Mele (Mele, 2013) conceptualizes the concept of weapons by pointing out, that "*a weapon can be [...] an abstract concept thereby not necessarily a material one*". He proposes the consideration of both the strategic dimension – intended damage and the specific selection of a sensitive target – as well as the legal dimension – context and purpose – of a cyber operation. The author defines these as the "*typical elements of a cyberweapon*". A similar premise is followed by Dewar (Dewar, 2017), who criticizes that "*the subjectivity and context-dependence [...] causes particular difficulties when categorizing cyber tools as weapons*" because "*all weapons are tools, but not all tools are weapons*". To resolve this, the author urges to "*look at more conceptual issues regarding the incidents*" and to evaluate the assumed motivations of an attacker. As a conclusion, he states that "*often the tool itself was not a digital device [...]. Techniques such as social engineering and phishing or the*

exploitation of unknown weaknesses in digital systems were the preferred tools". Orye and Maennel (Orye & Maennel, 2019) extend this assertion further by considering also the cognitive effects, which include "*sowing confusion, changing behavior, modifying trust, changing (public) opinion, manipulation*".

10.2.3 *Classifications Based on Normative Aspects*

An approach that analyzes the significance and impact of malicious cyber tools in international relations has been proposed by Stevens (Stevens, 2018). The author states that "*weapons can be understood as 'the violent materiality of the existential condition of uncertainty'*" and that cyberweapons "*shape a condition of marked uncertainty in the contemporary international order. Silent, invisible and potentially very effective, they are attractive to states and non-state actors seeking advantage in war and in peace*". This approach fundamentally questions the definition of cyberweapons in the context of global power and governance. An early attempt for regulation was undertaken with the extension of the Wassenaar Arrangement on Export Controls for Conventional Arms and Dual-Use Goods and Technologies (The Wassenaar Arrangement Secretariat, 2014). In 2013, the multilateral treaty added the item of "*intrusion software*" to its list of regulated goods, with the following definition: "*Software specially designed or modified to avoid detection by 'monitoring tools', or to defeat 'protective countermeasures', of a computer or network-capable device, and performing any of the following: (a) The extraction of data or information, from a computer or network-capable device, or the modification of system or user data; or (b) The modification of the standard execution path of a program or process in order to allow the execution of externally provided instructions*." This approach defines cyberweapons by their potential capacity for malicious impact on IT systems according to the effect-based classification.

10.2.4 *Classifications Based on the Comparison with Traditional Weapons and Weapons Technology*

Another attempt to define the nature of cyberweapons is based on the comparison with existing, well understood and examined weapon technologies. An early, but still important approach by Sommer and Brown (Sommer & Brown, 2011) reflects the features of a generic kinetic weapon. In their point of view, a weapon is "*a directed force – its release can be controlled, there is a reasonable forecast of the effects it will have, and it will not damage the user, his friends or innocent third parties*". The authors highlight that "*there is an important distinction between something that causes unpleasant or even deadly effects and a weapon*". They suggest evaluating cyber tools based on these characteristics, in addition to the required usage capabilities, the target "*inside knowledge*", whether and how fast the tool can be detected before and during deployment, how quickly counter measures can be applied, and whether it is possible to detect and identify the perpetrator. An attempt by Hatch (Hatch, 2018) states that the conditions for a cyberweapon are the system's fundamental design and initial consideration to "*act as a weapon*" and the "*capability to cause mass causalities at a single point in time and space*" like "*cyber operations that trigger a nuclear plant*

meltdown; open a dam above a populated area [...] or disable air traffic control services, resulting in airplane crashes".

10.2.5 *Classifications Based on Architectural Characteristics of Malicious Software*

A few proposals have been made that focus on the architectural characteristics of malicious software. One major approach is presented by Herr (Herr, 2014) with the "*PrEP framework*", signifying "*propagation, exploits and payload*" as the base components of any malware that are required for a cyberweapon: "*The propagation method defines how the code is delivered into a target system and the payload is the core executable code of the malware that determines its functionality and delivers its effects. The [...] exploit, allows both propagation and payload delivery by taking advantage of vulnerabilities in computer systems and their defensive measures*". The author also explicitly criticizes the intent-based perspective, because "*intent and perception are diffuse characteristics and difficult to judge"*. The distinction between the different components of a malicious tool aims to provide an in-depth view on the specific technical elements that define a cyberweapon. A critical "*red line*" is specified in particular by "*a payload designed to create destructive physical or digital effects*", whereas the authors state that cyberweapons "*create physical and digital effects"* and that *"defining [cyberweapons] without them creates unhelpful limitations*". Contrary to the authors criticism of "*effect and intend*", this still includes the assessment of the actual impact and the anticipation of an attacker's aims into the cyberweapons assessment. Besides the important technical perspective, the approach does not further exemplify how to measure the capabilities of the payload, nor does it provide a structured and uniform analysis method of the malware.

Another technical approach by Maathuis et. al. (Maathuis et al., 2016) defines three different components or layers of malware: The access layer to reach and enter a foreign IT system and circumvent any defense mechanism, the transport layer for the propagation within given IT infrastructures, and the payload layer for the effect. The authors state that a cyberweapon is "*a computer program created and/or used to alter or damage [...] a system in order to achieve (military) objectives against adversaries inside and/or outside cyberspace*". They suggest the assessment of different additional technical, tactical, and strategic aspects of the malware like the configurability and adjustability, the sophistication, and the technical knowledge about the intended target as well as the disruptive or destructive intent and the selection of a relevant target.

10.2.6 *Research Gap*

The presented selection of definitions and classifications of cyber weapons shows that most approaches utilize an assessment of the intent of the attacker and the purposed potential or actual effects of the digital payload. These are valuable approaches for agreements that focus primarily on the political level and on norms for state behavior in the cyberspace. However, they are not applicable in advance of a specific incident and their presumed intent will always be influenced by speculation, political and strategic considerations, and interests of various relevant actors. This highlights the necessity

for assessment measures that are applicable regardless of subjective considerations and before the tool is used.

10.3 Technical Features of Cyber Weapons

In order to develop such an assessment model, that is independent of the actual usage of the malicious software or speculations about its intentions, the following section discusses features of malware that can be measured or assessed independently, especially their technical particularities. From such a technical point of view, operational military weapon systems consist of a multitude of different interoperating parts, materials and the underlying technologies for their development and production. Stripping down complex weapon systems into their components is particularly necessary for trade and export regulations. Our analysis of distinguishable, measurable features of cyberweapons is therefore divided into the following sections, that reflect the different steps from development to deployment of weapons:

- Production and storage
- Availability and steps to full operational capacity
- Deployment and operation
- Impact and evolvement of effects
- Results, successions, and damage

Each of the sections will sum up the identified parameters in a separate table.

10.3.1 *Production and Storage*

From a technical perspective, cyberweapons are basically complex IT products which do not necessarily form a monolithic system, but often consist of independent, interchangeable parts for different purposes and stages of their deployment, as mentioned in the previous section. Such a modular design, which is often developed based on frameworks or platforms, allows attackers to execute a target-specific reconfiguration, compilation, or extension of cyberweapons as well as the integration of new features. Such modularity sustains the effectiveness and longevity of developed components, as they can be reused. Examples are routines for the automatic propagation of malicious tools within networks, code that loads target-specific assets after infection, or command and control infrastructures. The usage of extendable frameworks also enables attackers to learn from obtained code samples of other malware and to extend their tools accordingly. On the other hand, reusing the same tool may be an indicator for attributing attacks to their origins, prompting attackers to continuously adjust their developments.

A very distinctive aspect of cyberweapons is their sole effectiveness in a specific environment that is prone to the utilized malicious code, like e.g., a specific vulnerability or

exploit, that has been built into the cyberweapon. Whereas most parts of a cyberweapon are developed based on common knowledge that is also used and applied in civil and commercial IT security sectors, the unique core of each cyberweapon consists of the knowledge about the target's vulnerabilities and the specifically tailored code to exploit these weaknesses. This part is the essential element in the development of cyber arms and is the object of the ongoing cyber arms race. This reveals an important difference from other weapons technologies, where secrecy about information on the functionality and capability of all parts and technologies of a weapon is often crucial. However, the target-specific tailoring of a cyber weapon potentially also requires a target specific testing environment to evaluate, ensure, and adjust aspects of the weapon deployment such as automatic propagation or payload triggering.

In terms of the longevity of exploitable vulnerabilities, studies have shown the potentially enormous life span of up to nearly seven years until their detection and closure (Ablon & Bogart, 2017). This problem is regularly confirmed by data breach reports (Verizon, 2019), which show that most of the detected cyber incidents are based on already known vulnerabilities, which have not yet been patched in the targeted devices. Modular cyber weapons allow attackers to combine an existing payload with a vulnerability that matches the current target system. Therefore, the developed components of cyberweapons do not require any special maintenance in order to be reliable, apart from preventing other actors from gaining knowledge of these cyberweapon and its capabilities and establish safeguards for "*digital weapons arsenals*" in order to prevent any threats to the actor's own IT systems and the national IT security of states (Schulze & Reinhold, 2018). Table 10.1 summarizes these parameters.

P1	Long-life production perspective, modular, extensible, and interchangeable design and software architecture
P2	Developed and equipped with tailored malicious code for a specifically selected IT target and its vulnerabilities
P3	Quality assurance and quality management in dedicated testing facilities or environments
P4	Implementation of an update mechanism to combine existing malicious payload with current, state-of-the-art penetration tools and exploits
P5	Existence of secure vaults to store the malicious payload and prevent an unintended outbreak

Table 10.1: Parameters regarding the production and storage

10.3.2 *Availability and Steps for Full Operational Capability*

The military deployment of weapon systems is usually targeted against specifically selected objects, which is often associated with an adjustment to the environment of the target and its vulnerabilities. In the case of cyberweapons, this preparation requires extensive knowledge of the object, information that can often only be gathered through reconnaissance operations, which potentially require hacking of minor, upstream IT systems. This highlights that an effective deployment relies on the capacities to circumvent all security measures on the way towards the target, including all upstream

systems. Since these activities must remain hidden in order to prevent premature detection, the time required for deployment is also a decisive factor. Some military strategists argue that it is necessary or efficient to implant cyberweapons or to create hidden access possibilities in strategically relevant IT systems as a precautionary measure. The US Cyber Command extended this approach towards a "*persistent engagement*" (US-DOD, 2018a), that includes the permanent deployment of cyber tools within adversary networks, forcing them to constantly observe, secure, and adapt their systems.

Understanding the parameter of the availability of a cyberweapon as presented in table 10.2 is a question strongly connected to the specific operational context. This can be a state in which all necessary knowledge and tools have been gathered, but active penetration itself has not yet occurred. A different interpretation considers the strategic planning behind "*persistent engagement*"-like approaches. Here, availability is understood as a state in which an exploitable path to the target already exists and the payload can be or has already been introduced into the target system, ready to be triggered. This includes all infrastructures such as command and control servers, which must be ready for use.

P6	Implementation of tactical exploitation capabilities to reach the intended target through upstream systems and security measures
P7	Technical ability for a preliminarily deployment, long-lasting detection prevention, and later payload execution
P8	Development and implementation towards strategic goals and planning, including future conflicts

Table 10.2: Parameters regarding the availability and steps for full operational capability

10.3.3 *Deployment and Operation*

As the deployment of cyberweapons takes place in multiple, consecutively triggered steps, such tools should be considered using a shell model, like the basic three-layer approach proposed by Maathuis et al. mentioned above (Maathuis et al., 2016), or more complex models such as the cyber kill chain (Wrozek, 2017), a concept based on the work of Hutchins et al. (Hutchins et al., 2011). Drawing from those, each shell should contain its own tools and capacities, whereas the actual configuration and required infrastructure depend on two parameters: First, the intended effect and impact, which can be tailored either to a specific target or to a class of IT systems or products. Second, the decision to what extent a deployed tool should be capable to propagate autonomously and trigger its payload. The possible options range from a "*fire and forget*" approach with an automatic operation based on built-in rules to a manually operated approach with command-and-control infrastructures that allow direct human control of the deployment process. Especially the chosen propagation mechanisms may limit the measures available to prevent unintended effects. Limiting an automatic infection to intended targets can be difficult to control, since the behavior of a particular code depends on the conditions of the actually infected system, which are often difficult to predict, possibly resulting in incalculable effects. Even without automatic payload activation, a widespread infection raises concerns about which of the infected systems

are relevant to military goals. Even though these considerations are not dealt with in the context of this article, they are directly related to the "*human-in-the-loop*" debates that are highly controversial internationally in the field of lethal autonomous weapon systems (UNGGE, 2019) and the problems of meaningful human control (Cannellos & Haga, 2016). A measure to prevent or stop the unintended propagation can be an explicitly built-in so-called "*kill switch*", a function that offers the possibility to completely shut down the operation of the cyberweapon. However, based on technical examination of detected malware samples, these functions are rarely used, as they are relatively easy to detect by the defenders, which undermines their effectiveness (Symantec, 2017). With regard to the penetration of IT systems, any unauthorized attempt to access an IT system can potentially damage that system. The circumvention of security and defense measures or the concealment of access from logging mechanisms manipulates the regular behavior of the system. Depending on the skills and expertise of the attacker and the information available about the attacked systems, this can have unintended or unexpected effects, leading to operational disruptions or system failures. In view of the international humanitarian norm of protection of civilian systems, this requires an assessment of each intruded IT system, what programs it runs and what external purpose it serves. These aspects are presented in the following table 10.3.

P9	Ability to steer the propagation and payload activation that allows human interaction
P10	Implementation of a "kill switch" or similar mechanisms to immediately stop the further propagation and payload activation
P11	Technical ability to detect and control the penetration of unrelated systems, assess its functions and exclude them to prevent unintended harm

Table 10.3: Parameters regarding the deployment and operation

10.3.4 *Direct Impact and Effects*

The potential impact of cyberweapons can cover a broad spectrum, and the effective impact is strongly influenced by the weakness and vulnerability of the target which is reflected in the parameters of table 10.4. If a targeted system is not prone to a utilized vulnerability, has recently been updated and patched with security fixes, or has implemented strong IT security measures that detect and stop unusual system functions, a cyberweapon will either not be able to penetrate the target at all or will fail to launch its malicious payload. Other attack methods are based on the regular use of IT systems by overloading their processing capacity, which usually leads to their temporary shutdown. Although mitigation techniques exist to some extent (Osanaiye et al., 2016), these attacks are very effective and are conceptually more difficult to prevent (Lavrenovs, 2019). The different attack or infection approaches influence the possible reaction time of the defending actors regarding their chances of mitigating the effects and the malicious propagation, and thereby the effectiveness of the cyberweapon. A payload that has been secretly implanted into a target system limits the available defensive options, in contrast to cyberattacks that attempt to openly disrupt services. Once the payload has been triggered, the evolvement of its effects can vary widely, depending on the

configuration of the cyber weapon and the situation of the attacked system. It can range from a direct and contained impact on the targeted system (first level effects), impacts on connected IT systems that rely on certain services or functions of the targeted system (second level effects), to effects on other connected systems, either through propagation or chain effects (third level effects). The complete impact estimation contains a high potential for miscalculations or failures. The attacker needs to make assumptions about the target, its environment, dependencies, and the reaction of the attacked systems and actors while ensuring that the programming of the cyberweapon operates as expected and contains no errors. Nevertheless, a specific deployment may encounter unexpected conditions which can lead to a completely different effect or undesired effects.

The discussion on the participation of the China-based company Huawei in the construction of 5G mobile networks (Rupprecht et al., 2018) highlights another aspect. Concerns have been expressed that malicious code could be directly integrated into widespread small off-the-shelf components, possibly granting unauthorized access or waiting for a trigger signal. In such cases it is extremely complicated to distinguish between the legitimate host system and the malicious code, especially when backdoors are suspected to be hard-wired into the chip design.

P12	Time to react for a defender, range of possible defense measures
P13	Assured reliability, accuracy and containment of impact
P14	Degree of separation from any required host systems

Table 10.4: Parameters regarding the impact and evolvement of effects

10.3.5 *Overall Results, Successions and Leverage*

A comprehensive assessment of the possible overall effects of a cyber weapon, which is presented in table 10.5, is required for the authorization of its application in light of the rules of international law such as the UN Charter (Sander, 2019). For the current state of ongoing international debates, which have not yet been settled, the legal threshold is drawn at the point where the effects of a cyberweapon correlate with the "*use of force*" – usually interpreted as severe damage to objects or people – which is prohibited outside declared military conflicts (Kosseff, 2019). A complete evaluation must also include the aftermath such as the reaction of the attacked and third-party actors. If the utilized malicious code uses zero-day exploits or other methods of intrusion unknown to the public, its usage reveals this secret, allowing defenders to adjust their protective measures. It also provides any other witnessing or later analyzing party with knowledge to learn and adapt, as long as the vulnerability is not completely fixed. This can result in threats to the attacker's own systems, as demonstrated by the EternalBlue vulnerability, which – originally owned by the NSA – was used in the malware campaign NotPetya (ESET, 2018) that also caused economic damage to industrial facilities in the US (Maersk, 2017). As already mentioned, neither the built-in logic of a tool nor a human conductor can safely and ultimately decide whether the penetrated system is a valid military target or not. The risk of mistakes is especially present during the intrusion, since at this point the attacked system can only be analyzed "*from the outside*". The potential effects on uninvolved systems and the risks of maloperation highlight that

the actual effects caused by a cyberweapon can deviate considerably from its intention. A comprehensive assessment of such complex situations must therefore consider the following three dimensions to estimate the maximum possible effects:

- The time span for the unfolding of triggered effects and their evolvement on each affected system. This can range from immediate to delayed and restrained effects.
- The spatial dimension of the triggered effects, assessing the number of systems that may intentionally affected directly and indirectly as well as potentially unintentionally targeted lateral systems.
- The precision of the effects that can by triggered by the payload. This dimension needs to consider intended and unintended effects and can range from accurate, specific effects on a targeted system to maximum effects from "*brute force*" affecting all running and active services on a system or within a network.

P15	Potential for proliferation of know-how or the knowledge of vulnerabilities
P16	Time span, spatial dimension, and precision of the effects and the possible impact on directly, indirectly and potentially unintentionally affected systems, including self-harm

Table 10.5: Parameters regarding results, successions and damages

10.4 Assessment Model for Cyberweapons and Case-Study-based Evaluation

10.4.1 *How to assess cyber weapons*

The following section will propose an assessment model for cyberweapons that analyzes the capabilities of a given software and – given the intended application context of arms control – contrast the results with existing norms and regulations. The model cannot and does not aim to provide comparability between assessments – which would require any kind of scoring – but rather to present a structured method to assess specific features of a given software in order to provide a unified basis for the evaluation of its cyberweapon character. Therefore we propose a set of "cyberweapon indicators" as given in table 10.6, whereas each indicator is linked to multiple of the previously discussed technical parameters that relate to this indicator and connected with a possible range of expression. The indicator order follows the concept of the cyber kill chain (Wrozek, 2017), (Hutchins et al., 2011), an established method of malware life-cycle analysis that separates the different steps from the development of a malware to its deployment. Facilitating this order supports the unified assessment of the technical capabilities required for each step.

	Indicator	Range of expression / Assessment	Associated Parameters
I1	*Means of propagation*	Targeted and tailored measure vs. randomly spread approaches	P6, P8, P10
I2	*Autonomy of deployment and application*	Controllable and (re)configurable by human conductors vs. automatically decided by built-in rules	P7, P8, P9, P10, P14
I3	*Controllability and intervention measures*	Completely human decision on the triggering of the payload and the possibility of stopping its evolvement vs. Fire-and-Forget	P3, P5, P9, P13
I4	*Required infrastructures*	Degree of supporting infrastructures such as command and control systems, communication channels, data drop-off points	P1, P7, P8, P14
I5	*Quality of penetration measures*	Uniqueness and distribution of the exploited vulnerabilities and exploits code	P1, P2, P4, P12
I6	*Direct payload effect*	Type and degree of maximum impact to which the payload is intentionally programmed	P2, P4, P6, P12, P16
I7	*Unintended effects*	Measures of quality assurance and testing during the development phase, probability and measures for the handling of unexceptional situations over the full application process	P3, P5, P9, P10, P11, P13, P15, P16

Table 10.6: Indicators For Assessment Of Cyberweapons

The intended application of our proposed assessment model will stepwise assess each indicator by analyzing the associated parameters and the underlying questions regarding specific technical capabilities of the analyzed software. For the discussed context of arms control, this assessment will be performed by authorities like e.g. the German Federal Office of Economics and Export Control (BAFA - Bundesamt für Wirtschaft und Ausfuhrkontrolle) or in the U.S. Department of Commerce's Bureau of Industry and Security (BIS) that are entitled and authorized to review and grant or deny export requests based on national laws and regulations. As already implemented for other technologies and goods, companies requesting an export license are required by law to provide technical documentation, source code samples, or compiled binaries of their products and submit these to the authorities. Taking into account that these documents and required information are probably not complete, the assessment results for each

parameter can range between “yes”, “no”, “partially”, and “unknown”. This does provide neither a scoring nor binary answers, but taken together, its assessment allows to specify a position for each indicator within the provided range of expression. Although this leaves room for different considerations, the parameter assessments that focus on a specific capability and the question if a software contains it, provides in our opinion an appropriate degree of objectivity. Finally, the different indicator assessments in connection with the amount and distribution of “yes” and “partially” answers for the analyzed parameters provide the basis for a concluding decision on the cyberweapon character of the analyzed software. With regard to the intended application context, this decision will primarily be subject to legal regulations and political considerations. As usual for arms control and export regulation, the critical thresholds which technical capabilities are considered to manifest a weapon will likely differ for different states as long as no internationally binding norm or other treaties exist.

10.4.2 *Case-study based evaluation*

In order to evaluate the application of the identified indicators, the following section presents exemplary assessments, to find out whether the model applies to real-world cases. As there have been several incidents over the last years that have caused damage, we have chosen two “positive” cases that probably could be considered as cyber weapons and face these with one “negative” case, that is probably no cyberweapon in order to illustrate the contrast. With regard to the support decision character of our model, this is intended to be as an exemplification of its application, rather than a sufficient validation that would require systematic testing of cases that have been publicly classified as cyberweapons and cases that have not been labeled this way. This goes beyond the scope of this paper but is a task for future work. The intended use case of our model requires access to technical documentations and code samples, something that requires legal access possibilities for entitled authorities as these information are usually classified. To simulate this situation, we have chosen example cases of past incidents that have been provided with freely accessible analytical reports to test our assessment against the public assessment of the incident. The reports that we used have to focus on technical details of the malware such as reverse engineered and decompiled binaries, code samples, string analysis, comparison with known vulnerability and exploit databases, analysis of the propagation and communication capabilities. All technical information that we have taken into account should have been available to entitled arms control and export regulation authorities to this extent, even before the malware was used. We neither used knowledge of the malware outcome nor assumptions about the attacker’s intentions. Based on a meta-analysis of the selected reports, the evaluation will assess the cases by testing the indicators stepwise and analyzing each associated parameter. To circumvent the current lack of any internationally binding legal definition, we facilitated the broadly approved Tallinn Manual perspective (CCDCOE, 2013) for the final cyberweapon consideration in a slightly modified version where cyberweapons are tools that specifically contain the technical capability “[...] *of causing either (i) injury to, or death of, persons; or (ii) damage to, or destruction of objects*”. The following subsections will briefly present the cases, the reports we used, our assessment, and finally the conclusion. In order to maintain the readability of the text, we dispense with single code examples in this section, but reference to selected examples of technical details and further descriptions that we present in the Annex of this paper. In addition, we

list references and – if available – page numbers to the most relevant analytical findings and quotes of the reports to underline the technical foundation of our assessment. The detailed results for all indicators and assessed parameters are presented in table 10.7, where the assessment results are represented symbolically for a better overview with the following notation: "Yes" (●●), "No" (○○), "Partially" (●○) and "Unknown" (××).

10.4.3 *First positive Case: Stuxnet*

Although the Stuxnet incident dates back to the year 2010, it is presumably the best known and still the most thoroughly analyzed malware to date. It was discovered at the nuclear enrichment facility in Natanz, Islamic Republic of Iran and was used to achieve a beyond-the-normal wear of enrichment machines. Our assessment is primarily based on the analysis of Langers' "To kill a centrifuge" (Langner, 2013) and the highly technical Symantec reports on the initially detected (Falliere et al., 2011) as well as an earlier version of Stuxnet (McDonald et al., 2013). The authors conclude that Stuxnet has been tailor-made as it e.g., had been manipulating the supervisory control room software that "*appears to be a genuine development for the Natanz Fuel Enrichment Plant*" (Langner, 2013, p.8) and contained exploits for its specific vulnerabilities [Annex A.1] to hide its activities as well as to intercept and manipulate the Step7 dubbed control of the industrial programmable logic controller (PLC) [Annex A.2] which regulates the actual industrial process. In addition, the code contained information on the specific configuration of the industrial hardware in this facility and "*infects [...] controllers with a matching configuration*" (Langner, 2013, p.8), [Annex A.3]. Stuxnet contained at least two different attack payloads, one that manipulated the rotor speed in centrifuges (Langner, 2013, p.10ff) and an earlier, yet much more dangerous version that can create overpressure in the enrichment devices (Langner, 2013, p.5ff; McDonald et al., 2013, p.9ff): "*[it] contains an alternative attack strategy, closing valves within the uranium enrichment facility at Natanz, Iran, which would have caused serious damage to the centrifuges and uranium enrichment system as a whole*" (McDonald et al., 2013, p.1). Regarding the control of the attack, Stuxnet was able to develop a communication channel despite the air-gapped system by facilitating infection and propagation methods that allowed to transfer information "*by compromising mobile computers of contractors who enjoy legitimate physical access to the target environment*" (Langner, 2013, p.22; [Annex A.4]. In contrast, the authors conclude that "*there is no logic implemented in the malware which could actively disable the malicious code on infected controllers*" (Langner, 2013, p.18). In line with these examples, the evaluation of all parameters draws a picture of a project that contains the capabilities to reach, attack, manipulate, and even destroy a specific IT system. The attacker exploited multiple zero-day vulnerabilities and had made significant efforts to avoid unintended side-effects or the detection of the attack and invested considerable know-how to establish a communication channel despite the limited direct controllability. Besides some several high-class exploits, Stuxnet contained no off-the-shelf utilities or code. Taking all of this into account, this assessment confirms the conclusion that Stuxnet has to be considered a cyberweapon.

10.4.4 *Second positive Case: TRISIS/TRITON*

The second example TRISIS/TRITON has been detected in 2017 in a petrochemical plant in Saudi Arabia. It presumably was manually injected and able to manipulate the Schneider Electric's Triconex safety instrumented system (SIS) that is responsible for reacting on critical operation incidents to deactivate the fail-safe operation of the industrial facility. For our analysis we used three reports from Dragos (Dragos Inc., 2017), FireEye (Johnson et al., 2017), and the US National Cybersecurity and Communications Integration Center (NCCIC, 2019). TRISIS code was designed to target a specific facility, identified by a SIS configuration that the malware was designed for: "*[As] each SIS is unique and to understand process implications would require specific knowledge of the process*" (Dragos Inc., 2017, p.3). The malware required a manual injection to the facilities network [Annex B.1] and exploited a vulnerability of a specific version of the Triconex system [Annex B.2]. The malware contained code to perform different alternative attack methods (Johnson et al., 2017, p.4) to manipulate or deactivate the SIS system "*that collectively would degrade industrial processes, or worse. Were both the process and the safety systems to be degraded simultaneously, persons, property, and/or the environment could suffer physical harm"* (NCCIC, 2019, p.18, p.12ff). TRISIS code was written in Python and based on structure of the code, the possibilities to extend it with additional scripts it "*represents a facilitating capability or framework for the actual ladder logic change that has the potential [... to] be repurposed to deliver alternative payloads to either deliver different logic files (the external binaries uploaded by TRISIS to the target SIS) or to utilize differently embedded binaries to target different SIS types entirely."* (Dragos Inc., 2017, p. 13). As TRISIS is built to be operated fully manually via hardcoded communication channels (McDonald et al., 2013, p.12) the code contained "*anti-forensics technique to hide the presence of the attacker code on the Triconex controller"* [Annex B.3]. Regarding the complex and diverse infection, the capabilities for the evasion of security measures and the code injection process, the reports conclude that TRISIS development needed highly qualified, well-resourced adversaries with lengthy timelines (McDonald et al., 2013, p.8). This assessment of TRISIS shows that it contains capabilities to attack a strategically selected goal with methods for a rather long-term access and manipulation capabilities which could cause serious damage. This suggests the conclusion that TRISIS must also be considered a cyber weapon.

10.4.5 *Negative Case: Emotet*

The following subsection will present an incident, that - although it has caused serious damage - cannot be considered to be cyberweapon according to our approach. This negative case example should illustrate the relevance of certain indicators and specific technical capabilities for our assessment and classification approach. The example we have chosen is Emotet, a trojan malware that was first detected in 2014 (Axel, 2020). During its lifetime, the malware has been changed a lot and often been used as an preliminary infection step for multiple different malware campaigns that afterwards loaded additional payload code. According to the US-CERT, the malware is "*among the most costly and destructive malware affecting state, local, tribal, and territorial (SLTT) governments, and the private and public sectors*" (CISA, 2020). At least one

case is reported, where an infection directly impaired hardware by overheating (Gatlan, 2020). With regard to this threat, the Emotet infrastructure has been took down in 2021 by an international coordinated action, led by EUROPOL (Europol et al., 2020). For the technical assessment we used reports from Bromium (Bromium, 2019), Malwarebytes (Malwarebytes Blog, 2021), PaloAlto Networks (Duncan, 2021) and the US Cybersecurity and Infrastructure Security Agency (CISA) (CISA, 2020). According to these reports, Emotet is a flexible malware that facilitates a very "*effective combination of persistence and network propagation*" (Malwarebytes Blog, 2021). Over the years, the malware has emerged to an actual business, where "*the primary source of revenue for its operators may be through selling access to its botnet infrastructure to other malware operators, instead of directly monetizing stolen financial information*" (Bromium, 2019, p.3). Therefore, Emotet has developed to a broad toolbox that utilizes different phishing and watering hole infection methods: "*Emotet uses different techniques to distribute these [infected] Word documents. The malspam may contain an attached Microsoft Word document or have an attached ZIP archive containing the Word document. [...] Some emails distributing Emotet do not have any attachments. Instead, they contain a link to download the Word document. In previous years, malspam pushing Emotet has also used PDF attachments with embedded links to deliver these Emotet Word documents*" (Duncan, 2021). In addition, Emotet established a complex botnet infrastructure that is used, among other things, to deliver different payloads or download additional code from a set of an extendable set of modules for specific data grabbing and exfiltration tasks (CISA, 2020). Although Emotet has been used for different campaigns, from simple ransomware attacks to attacks on more selected targets and state institutions like parts of the Lithuanian Government (Staff, 2020), its technical capabilities are focused on broadly executed phishing activities [Annex C.1], with a high degree of automated operation and a broad impact. Even if targeted campaigns had facilitated custom target-relevant phishing emails or selected groups of email recipients, its technical capabilities are rather not built to be tailored made for a selected, certain ICT system [Annex C.2]. Besides the already mentioned different phishing methods of infected documents, this is also reflected in capabilities intended for spreading within networks: "*Once Trojan.Emotet has infected a networked machine, it will propagate by enumerating network resources and write to share drives, as well as brute force user accounts. Infected machines attempt to spread Emotet laterally via brute forcing of domain credentials, as well as externally via its built-in spam module*" (Malwarebytes Blog, 2021). In addition, Emotet did neither use zero-day exploits nor any unique target information like running service and its software versions and configuration, but rather exploited known vulnerabilities (CISA, 2020) of commonly used operating and office software with an "hit as much as you can" mentality and is even capable of injecting other malware into infected systems on a malware-as-a-service basis [Annex C.3]. These assessments lead to the conclusion, that on the one hand Emotet contains the capabilities for causing damage, but on the other hand lacks a specific target selection and tailoring as well as a manual steering of the attack, the payload delivery and its triggering. The code does not contain dedicated measures to prevent unintended proliferation and effects. Beside its indisputable destructive and dangerous nature, with regard to our assessment model, Emotet cannot be considered a cyberweapon and rather reminds of a tool for "digital vandalism" in the sense that it indiscriminately damages the things within its reach. This assessment is in line with the majority of Emotet's public perception.

10.4.6 *Evaluation of results*

Given the results of the presented examples, our proposed assessment model supports the existing interpretation of the incidents in all three cases. Both positive case studies have almost exclusively assessed parameters answered with “yes” or “partially”. In addition, these malware examples had been developed to damage a specific target and dedicated hardware over a long timeframe and utilized a diverse range of techniques for infection, propagation, or payload deployment. These can be considered the core technical features of a possible cyberweapons. Although having the capability for destructive effects, the negative malware example does not fulfill these characteristics. This presumably also applies to the aforementioned NotPetya incident, which created commercial damage but also lack these features. They too had been developed to cause as much damage as possible by utilizing self-propagation mechanisms as well as quickly exploiting a zero-day vulnerability, which was widespread at the time, for its ransomware and disrupting payload. The example assessments also showed that an assessment based on technical capabilities is practical applicable and provides a valid basis for a conclusion on the cyberweapons character of a software.

Indicator	Case 1: Stuxnet		Case 2: TRISIS/TRITON		Case 3: Emotet	
	Evaluation	Assessment	Evaluation	Assessment	Evaluation	Assessment
Means of propagation	• Automatic, covered spreading over intruded networks • Propagation until a specific network (target) reached • Payload only activated within target • Injection, communication and control probably over an air gap or manipulated hardware • Specifically designed for Siemens SCADA product line “Step 7” and industrial hardware devices	P6: ●● P8: ●○ P10: ●●	• Potentially manual infection • Several different modules for privilege escalation and access • Specifically designed to infect Schneider Electric’s “Triconex 3008 Safety Instrumented System” (SIS) controllers • Capability to manipulate failsafe behavior of industrial facility could have been used to force drastic damages with additional malware • Manual payload triggering and stopping	P6: ●● P8: ●● P10: ●●	• Multiple spreading mechanism, but following rather an “A lot helps a lot” approach for maximum proliferation • Capabilities to spread over networks quickly and establish backdoors for payload • No kill switch but similar IT security measures could be established temporarily by IT experts via exploiting a bug in Emotet	P6: ●○ P8: ○○ P10: ○○

Indicator	Case 1: Stuxnet		Case 2: TRISIS/TRITON		Case 3: Emotet	
	Evaluation	Assessment	Evaluation	Assessment	Evaluation	Assessment
Autonomy of deployment and application	• Intrusion, detection-prevention and propagation via built-in automatic routines for different Microsoft operation systems and Siemens software • High degree of concealment mechanisms • Monitoring, control, and software updates via different C2 servers and ad hoc P2P mechanism • Payload activation automatic based on built-in routines to detect the intended target, anticipating the air gap of the target	P7: ●● P8: ●● P9: ●○ P10: ●● P14: ●○	• Manual deployment towards a specific system • Manual operation within target network • Custom made infection routines and customized process manipulating payload for a specific industrial facility and its SCADA architecture • Continuously adjusted concealment mechanisms	P7: ●● P8: ●● P9: ●● P10: ●● P14: ●○	• Broad, automated deployment • No manual operation • Capabilities for hit-and-run campaigns with broad infection and payload propagation • Later additional payload download possible but no manual steering on single infected systems • Common obfuscation and encryption of malware files to prevent automated AV countermeasures • No kill switch • Botnet infrastructure for data exfiltration and additional payload provision	P7: ●○ P8: ○○ P9: ●○ P10: ○○ P14: ●○
Controllability and intervention measures	• Intended, but faulty automatic disablement of propagation via time settings of infected system • No dedicated "kill switch" • Probably a dedicated testing facility of industrial target • Limited communication with C2 servers, mostly built-in, but updatable	P3: ●● P5: ×× P9: ●○ P13: ●●	• Direct access to infected system and human controlled operation • Payload tailored based on situational conditions • Probably a dedicated testing facility of industrial target • Payload tested on infected device before finally deployed	P3: ●● P5: ×× P9: ●● P13: ●●	• Relatively "open" infra-structure meant to be operated by third parties • Custom payload possible, but not for single infected systems • "Targeted" only in the sense of dedicated email recipients and tailored phishing emails • No proliferation containment	P3: ×× P5: ○○ P9: ●○ P13: ○○
Required infrastructures	• Malware itself independent from C2 infrastructures • Communication and data exchange channels via different, separated measures	P1: ●● P7: ●○ P8: ●● P14: ●○	• C2 infrastructures for deployment and operation, Hardcoded DNS servers • Communication and data exchange channels via different measures • Extendable via loadable code modules	P1: ●● P7: ●● P8: ●● P14: ●○	• Modular software with different infection and persistence methods • Custom payload with option for later download • C2 botnet infrastructure for payload provision and data extraction	P1: ●● P7: ●○ P8: ○○ P14: ●○

Indicator	Case 1: Stuxnet		Case 2: TRISIS/TRITON		Case 3: Emotet	
	Evaluation	Assessment	Evaluation	Assessment	Evaluation	Assessment
Quality of penetration measures	• Different modules and measures for penetration and propagation • Multiple 0day exploits • Bridging the air gap • Redundant measures for application life cycle (infection, data drop off, communication) supporting the autonomous propagation and infection over different systems • Built-in extensive knowledge of target environment and vulnerability	P1: ●● P2: ●● P4: ●● P12: ●○	• Extendable, continuously refactored Framework architecture • Payload independent from interchangeable initial infection and persistence capabilities • Payload injection and operational code tailored for specific devices • Built-in extensive knowledge of target environment and vulnerability • If combined with harmful payload, immediate physical effects possible	P1: ●● P2: ●● P4: ●● P12: ○○	• Continuously extended software basis • Exchangeable infection, propagation and persistence methods • Independent payload • Not developed to reach and infect single target systems and exploit their specific vulnerabilities • Developed for broad, quick and effective infections and backdoor establishment • No zero-day exploits • Optional immediate payload triggering	P1: ●● P2: ○○ P4: ●○ P12: ●○
Direct payload effect	• Payload explicitly developed for a specific software version and production line of industrial hardware • Interchangeable Payload • Propagation mechanism developed to reach a specific target via multiple, different measures • Different measures for direct impact from direct immediate harm (v0.5) to slow sabotage (v1.0) • No direct defending possibility, but system shutdown	P2: ●● P4: ●● P6: ●● P12: ●○	• Payload explicitly developed for a specific software version and production line of industrial hardware • Payload interchangeable • By manipulating the failsafe behavior of the facility, direct harmful impact with additional malware possible without defending or mitigating possibilities • Manual operation allows to prevent collateral infections and payload deployment	P2: ●● P4: ●● P6: ●● P12: ○○	• Optional immediate payload triggering • Payload interchangeable and option for later download after backdoor established • Third party payload injection possible • No manual operation • No zero-day exploits used • Multiple spreading mechanism, following rather a “A lot helps a lot” approach for maximum proliferation • No actively harming payloads known	P2: ○○ P4: ●○ P6: ●○ P12: ●○
Unintended effects	• Presumably a high level of diligence by testing in a dedicated testing facility and replacement of Stuxnet v0.5 payload • Automatic propagation, but malicious payload triggered only on the target system • Integrated, though faulty “kill switch” • Precise impact • Potential spread of zero-day exploits	P3: ●● P5: ×× P9: ●● P10: ●● P11: ●● P13: ●○ P15: ●● P16: ●●	• Presumably a high level of diligence by testing in dedicated testing facility • Manual operation based on direct feedback prevented deployment and payload errors • Proliferation of new, highly critical attack vectors for SCADA systems • Uncalculatable destructive effects if combined with destructive malware	P3: ●● P5: ×× P9: ●● P10: ●● P11: ●● P13: ●● P15: ●● P16: ●●	• Automated, uncontainable propagation and infection • “Targeted” only in the sense of dedicated email recipients and tailored phishing emails • Optional automated payload triggering • No manual steering of payload triggering on single infected systems • Third party payload possible • No kill switch • No zero-day exploits	P3: ×× P5: ○○ P9: ●○ P10: ○○ P11: ○○ P13: ○○ P15: ○○ P16: ○○
Legend: The assessment results are symbolically represented with the following notation: “Yes” (●●), “No” (○○), “Partially” (●○) and “Unknown” (××)						

Table 10.7: Detailed Evaluation of Selected Case Studies

10.5 Conclusion and Future Work

10.5.1 *The Technical Assessment of Cyber Weapons*

Our research aimed to develop an assessment method for identifying cyber weapons within the complex and diverse landscape of malicious software, based on features that are determinable without an assessment of their application context or an already performed usage. Our analysis shows that this is possible based on existing technical parameters that can be collected, tracked, or counted, regardless of the prior usage of the malicious tool and independently of speculations about its intent. The individual range of our proposed assessment indicators underlines the fact that it is possible to identify tools which are being developed to get weaponized – thus constituting cyberweapons. With regard to the requirement of available technical documentations and code samples, the proposed assessment model can provide a valuable contribution to the regulation of such tools, like for the implementation of arms control and non-proliferation treaties.

10.5.2 *Limitations*

The assessment model with the proposed list of indicators does not claim to be exhaustive. It is rather intended to provide a standardized and unified procedure to determine if a specific malware can be considered a cyberweapon. In addition, the indicators can be utilized to cluster specific weaponizable functionalities of malware that characterize such weaponizable tools. Such a generalization, that considers a broad range of parameters, cannot provide "*sharp edges*" as dual-use aspects or incomplete information will influence decisions. The proposed approach is therefore optimized as a decision support for case-by-case assessments.This reflects also the limited area of application, as detailed technical documents, code samples or other technical information on the software are necessary. Therefore, the legal and institutional foundation to request and assess this information, e.g. as part of export control regimes or in the context of a vulnerability equity process (Schulze, 2019), as well as their sensitive and probably secret nature needs to be considered when conducting our proposed assessment. The completeness of the available information also directly influences the amount and the certainty of assessable parameters. Nevertheless, these specific requirements and the political will are given for the intended context of arms control and its application via entitled authorities that are legally allowed to request the required technical details.

10.5.3 *Further Research*

The indicators and parameters identified can provide applicable measures for evaluating the cyberweapon character of malicious cyber tools. In addition to case-by-case decisions, they can also be used to cluster existing malware based on their technological approaches and capabilities. Further research should explore the development of a deterministic indication algorithm that combines the indications to weighted numerical values in order to compare different tools as well as to establish decision thresholds.

In addition, a systematic study of more past incidents could support this refinement alongside a validation of the indicators as well as a possible identification of edge cases that need to be considered. Besides the task of the identification of cyberweapons, the analysis shows that the risks of unintended effects are high and depend on many aspects of the target system, some of which are difficult to assess. Further research can refine the definition of minimum considerations and implementation principles that help to minimize the risk of unintended effects, in line with international humanitarian law and its prohibition to attack "*objects indispensable to the survival of the population*" (Gisel & Rodenhäuser, 2019). Finally, the ongoing militarization of cyberspace, with its consequences for international security, require a substantial, non-commercially motivated involvement of the computer science and a commitment to political issues, as many political challenges of cyberwar and its prevention have a deep rooting in technical details. This affects the development of technical measures for cyber arms control and its non-proliferation, the assessment of cyberattack methods, or the question how military cyber activities could follow international human rights rules, such as the distinction between civilian and military objects in the cyberspace. An understanding of these technical challenges, the stronger cooperation between computer science and politics, and the "translation" between these domains may pave the necessary way towards a stable and secure global cyberspace.

10.6 Annex

This Annex presents selected examples of technical details regarding the assessment of the three selected cases from section 10.4.2

10.6.1 *Stuxnet*

1. List of exploited zero-day vulnerabilities for all detected Stuxnet versions
   - MS09-025
   - CVE-2010-2568
   - CVE-2010-2772
   - CVE-2012-3015
   - CVE-2008-4250
   - CVE-2010-2729
   - CVE-2010-2743
   - CVE-2010-3888

2. PLC-Step7 Communication Manipulation

   "*The Step7 software uses a library file called s7otbxdx.dll to perform the actual communication with the PLC. The Step7 program calls different routines in this .dll file when it wants to access the PLC. For example, if a block of code is to be read from the PLC using Step7, the routine s7blk_read is called. The code in s7otbxdx.dll accesses the PLC, reads the code, and passes it back to the Step7 program. Stuxnet [...] renames the original s7otbxdx.dll file to s7otbxsx.dll. It then replaces the original .dll file with its own version. Stuxnet can now intercept*

any call that is made to access the PLC from any software package" (Falliere et al., 2011, p.37)

3. Target checking and selection process

   Stuxnet searches for an industrial plant from Siemens with a specific hardware configuration by searching in the code for "*symbol labels [that] loosely follow the ANSI/ISA S5.1 Instrumentation Symbols and Identification standard used in Piping and Instrumentation Diagrams*" (McDonald et al., 2013, p.6). Each PLC is identified by a fingerprint label following this format: "*<delimiter><FunctionIdentifier>*
   <delimiter><CascadeModule><delimiter><CascadeNumber>
   <DeviceNumber>" (McDonald et al., 2013, p.6). Since the concrete PLC configuration of an industrial plant is unique, Stuxnet checked the amount and type of all PLCs it detected and compared this against a built-in list of PLC fingerprints to identify a specific industrial facility.

4. Jumping the Air gap via removable drive infections

   "*Stuxnet will copy itself and its supporting files to available removable drives any time a removable drive is inserted, and has the ability to do so if specifically instructed" (Falliere et al., 2011, p.29ff), thus ex*ploiting the LNK vulnerability CVE-2010-2568. "*Stuxnet will first verify it is running within services.exe, and determines which version of Windows it is running on. Next, it creates a new hidden window with the class name 'AFX64c313' that waits for a removable drive to be inserted (via the WM_DEVICECHANGE message), verifies it contains a logical volume (has a type of DBT_DEVTYP_VOLUME), and is a removable drive (has a drive type of DEVICE_REMOVABLE).*" After checking if the drive is suitable, "*.lnk files are created using Resource 240 as a template and four are needed as each specifically targets one or more different versions of Windows including Windows 2000, Windows XP, Windows Server 2003, Windows Vista, and Windows 7. The .lnk files contain an exploit that will automatically execute ~WTR4141.tmp when simply viewing the folder.*" (Falliere et al., 2011, p.29ff) to inject Stuxnet into the system processing, allowing its hidden operations.

10.6.2 *TRISIS/TRITON*

1. *Target discrimination and spear headed design*

   "TRISIS is a Stage 2 ICS Attack capability, as defined by the ICS Cyber Kill Chain (...). Given its design and assessed use, TRISIS has no role or applicability to IT environments and is a focused ICS effects tool. As a result, TRISIS' use and deployment requires that an adversary has already achieved success in Stage 1 of the ICS Cyber Kill Chain" – Identifying and gaining access to a system able to communicate with target SIS – "and either compromised the business IT network or has identified an alternative means of accessing the ICS network. Once in position, the adversary can deploy TRISIS on its target: an SIS device." *(Dragos Inc., 2017, p.9)*

2. Exploited vulnerabilities for privilege escalation

Triton leverages a "*previously-unknown vulnerability affecting Tricon MP3008 firmware versions 10.0–10.4 [that] allows an insecurely-written system call to be exploited to achieve an arbitrary 2-byte write primitive, which is then used to gain supervisor privileges*." Regarding the output addresses of the exploited system call "*No checking is performed (...) to ensure the pointers do not refer to the firmware region or other protected areas. This allows for data to be written to normally immutable and privileged regions.*" (NCCIC, 2018, p.15-16)

3. Anti-forensic and evasion techniques

   As Triconex and the SIS systems are highly safety-critical, they contain numerous fail-safe techniques, like checksum comparisons to ensure the validity of the code. Triton contained a dedicated module *crc.py within its* loaded *library.zip* of compiled Python modules that "*implements or imports a number of standard Cyclic Redundancy Check (CRC) functions*" (Dragos Inc., 2017, p.7) that are used to patch a "*specific RAM/ROM consistency check*" in order to "*prevent a fault from occurring when the firmware region does not match the ROM image that was loaded. Without patching this check, the injector would not be able to write the payload into the firmware region or modify the jump table to point to it without faulting the device.*" (Dragos Inc., 2017, p.14). Additionally, "*after payload files were inserted into memory on the Triconex controller, the script initiated a countdown, periodically checking the status of the controller. If an error was detected, the communication library's method SafeAppendProgramMod attempted to reset the controller to the previous state using a TriStation protocol command. If this failed, trilog.exe attempted to write a small 'dummy' program to memory. We assess that this was an anti-forensics technique to hide the presence of the attacker code on the Triconex controller" (Johnson et al., 2017).*

10.6.3 *Emotet*

1. Phishing and data breaching variations

   "*Emotet uses five known spreader modules: NetPass.exe, WebBrowserPassView, Mail PassView, Outlook scraper, and a credential enumerator" (CISA, 2020). These* different tools can be used independently by loading different payload files into the memory once the victim is infected and information about the system are sent back to the C2 servers. The payload ranges from functions to collect passwords and user credentials from infected systems and external drives, read email addresses from Outlook accounts, send phishing mails, collect passwords and credentials from different web browser storage files, fetch "*passwords and account details for various email clients such as Microsoft Outlook, Windows Mail, Mozilla Thunderbird, Hotmail, Yahoo! Mail, and Gmail" (CISA, 2020) and* query network resources for further vulnerable systems.

2. Variations of initial infection mechanisms

   *Emotet*s first infection step is the spreading via different spam campaigns that lure the victim into downloading the malware. *"The email content may have a malicious link leading the victim to the Emotet downloader, or in other cases the downloader is delivered as the email attachment. We have seen MS Office*

Word documents, Excel spreadsheets, PDFs, JavaScript, and even password-protected ZIP files as the attachment. The most highly evolved spamming method, which appeared in recent months, is when the malicious object is inserted into a legitimate email conversation thread" (Nagy, 2019). In each case, malicious code is loaded from a range of different C&C servers either via direct download, VBA macros, or MS Windows shell functions.

3. Malware-as-a-Service capabilities

   Beside the malware's own payload files, "*Emotet has the ability to install other malware and to infect the machine with it. There are examples where it has distributed other banking trojans including Qbot, Dridex, Ursnif/Gozi, Gootkit, IcedID, AZORult and Trickbot and then ransomware such as Ryuk, BitPaymer or MegaCortex. In cases where additional malware is delivered besides the modules, the executeFlag in the response is set to 0x03, leading the delivered malware to the 'C:\ProgramData' folder with a randomly generated name. I have seen a downloaded Ursnif variant with a list of the most common latest modules. It injected control.exe under the 'C:\Windows\System32' directory, which further injected code into explorer.exe. It copied itself to the '%APPDATA%\Microsoft\[random]' folder and set the AutoRun registry to gain persistence*" (Nagy, 2019).

11

SPOTTING THE BEAR: CREDIBLE ATTRIBUTION AND RUSSIAN OPERATIONS IN CYBERSPACE

ABSTRACT How do we know who is behind a cyberattack? What are the tools and techniques that could help to identify the hackers who have conducted a cyber-operation? And why is credible attribution in the case of cyberattacks carried out or masterminded by Russia so challenging? These are the questions which this chapter aims to address in detail. However, before examining the technical, intelligence and geopolitical aspects of attribution, this chapter will first explain what attribution is and why it is important in the domain of cybersecurity

ORIGINAL PUBLICATION *Herpig, S., & Reinhold, T. (2018, October). Spotting the Bear: Credible Attribution and Russian Operations in Cyberspace. In N. Popescu & S. Secrieru (Eds.),* Hacks, leaks and disruptions: Russian cyber strategies *(pp. 33–42, Vol. 148). European Union Institute for Security Studies (EUISS). https://www.jstor.org/stable/resrep21140.7*

11.1 ATTRIBUTION: WHAT IT IS AND WHAT FOR?

The term 'attribution' chiefly refers to a concept in international law that describes the process of identifying an attack or operation against a state. It is widely used in debates about the norms and rules of state behavior and how they apply to cyberspace. Strictly speaking, attribution is usually used in the context of armed attacks and considered as one of the main legal requirements for the right of states to resort to force in self-defense under article 51 of the UN Charter. Any state action that refers to this article has to be justified by the credible identification of the origins of an attack. Only by supplying such evidence, are states deemed to have the 'inherent right of individual or collective self-defense' and can therefore take steps towards an appropriate response to stop these threats. Attribution is also used to convince the state's own government, public, and transnational partner organizations about the origin of a cyberattack. The threshold to convince these parties might be lower than the one required to trigger article 51.

In the case of national attribution policymakers and the executive have to be convinced about the origin of an attack in order to legitimize the use of offensive countermeasures. This form of assessment can rely on all technical, geopolitical and intelligence data that is available. Transnational attribution refers to convincing allies, bilaterally or as a whole (e. g. through NATO), about the validity of the attribution. This is essential in order to obtain political, diplomatic or other kind of support from allies. It might not be possible to explicitly use or cite certain intelligence and technical data to convince them, as such data may be too highly classified. Naturally, this might limit the credibility of

the attribution analysis. Then there is public attribution which refers to convincing the public with the attribution assessment. Having public backing in a democratic country allows the government to choose from a wider range of policy options. Credible public attribution enables the government to take certain steps, such as expelling diplomats or implementing economic sanctions, or at least facilitates such a response. Unlike in the case of transnational attribution (for which purpose allies might partially share sensitive information), most intelligence information and certain technical data cannot be shared with the public due to its classified and highly confidential nature. And this very limitation might undermine the credibility of public attribution.

11.2 Attributing a Cyber Operation

11.2.1 *Technical aspects of attribution*

From a technical point of view, measures for attribution need to be articulated in at least two dimensions. The first dimension distinguishes between those measures that can be established in the domestic IT systems and networks of a state ('inner scope measures') and those that need to exist or be built up in foreign IT systems ('outer scope measures'). The second dimension is composed of preventive and reactive measures. Preventive measures constantly observe, collect and store data that could be used to identify an attack that is detected or noticed at a later point in time. Reactive measures can be used to 'mark and track down an attacker' during an ongoing operation.

Inner and outer scope measures are both crucial to establishing credible attribution. Therefore, each state would need to have implemented its own set of data-gathering measures and/or allow defenders to trail attackers through their systems. This is clearly a challenge in the realm of international relations. Defining international standards for data-gathering measures, cooperation guidelines, information sharing and known communication channels would go a long way towards addressing this challenge and creating a common process to enable an international response. A first step towards such international cooperation has been made by the Budapest Convention on Cybercrime[1] which became effective in 2004. Its agreements however do not apply to norms of state behavior like espionage or military cyber activities and its practical aspects still need additional and more simplified cooperation measures. The United Nations had been moving in a similar direction until its group of government experts (UN GGE) failed to reach consensus in 2017.[2] Microsoft's private initiative, the 'Digital Geneva Convention',[3] is currently the most recent development in this area.

[1] See Council of Europe, Details of Treaty no.185, Convention on Cybercrime, Budapest, November 23, 2001 (EU-Council, 2001)

[2] Adam Segal, "The Development of Cyber Norms at the United Nations Ends in Deadlock. Now What?," Blog post, Council on Foreign Relations (CFR) (Segal, 2017)

[3] Brad Smith, "The Need for a Digital Geneva Convention," (Smith, 2017)

Monitoring and logging

A first and essential preventive approach involves technical measures that monitor access to IT systems, the connections and data transferred between them as well as user-performed operations like creating, editing, copying or deleting files. The level of detail of the collected data and the retention period[4] play a crucial role because in the absence of these elements investigations of attacks may not be pursued effectively. On the other hand, however, these same elements constitute a sensitive area in terms of data privacy. This issue has recently been debated in Germany, when allegedly Russian attackers broke into the Federal Foreign Office and undermined security mechanisms set in place by the secure government network.[5] The data storage is considered sufficient when logged information covers the entire attack within a specific system. This enables the defender to consolidate a detailed timeline about the attacker's actions, what the origin of the attack was, what data has been extracted and to which location the stolen data has been transferred.[6] Additionally, the logged information needs to be stored in a secure and tamperproof way to prevent attackers from erasing their digital footprints.

Computer forensics

When an attack has been detected, there is a range of possible reactive measures that can help in identifying the attacker. Besides analyzing the collected data to trace the attacker's operations history, other measures are the search and collection of software or software fragments that attackers have left on the compromised system to perform their unauthorized activities. These tools are often handcrafted and form part of larger tool sets. They are frequently reused for different operations over a period of several years. A software code analysis of these tools can be instrumental in detecting similarities and establishing connections with former incidents. This ranges from the language, geolocation or working hours uncovered by this form of assessment to code fragments and linked IT-infrastructures, such as email addresses and device IPs. This analysis therefore helps to identify familiar tactics and hacking approaches, linking them to known malicious actors.

Passive tracking

Passive tracking gives the defender additional information to potentially identify the attacker. While the attack is still in progress valuable information and evidence can be collected if the defender is able to observe the attacker's operations. This can be achieved by luring the attacker with so-called 'honeypots': systems or flaws that are easy to exploit and therefore will probably be targeted by the attacker. If the attacker takes the

[4]The retention period for stored information can be a critical aspect because sometimes attackers break into systems and create backdoors but then stay silent over a long period of time. When the attack is carried out, the log files only contain data about the 'strike command' but not necessarily the more significant information about the break-in itself.

[5]See Dana Heide, „Will der Bund die Cybersicherheit erhöhen, muss er den Datenschutz opfern," *Handelsblatt* (Heide, 2018)

[6]This is just an example. Usually log files can contain a lot more data and specific information on tampered data, modified executable files etc.

bait, the honeypot enables the defender to monitor all the attacker's actions.[7] A similar approach is the presentation of manipulated documents, relevant data or information that an attacker is potentially looking for and which contain malicious code, specific digital fingerprints or slightly manipulated information that can later be used to identify the data when it resurfaces.[8] These so-called 'beacons' might also send back the IP address of the systems to which they have been transferred, which could reveal the original location of the attacker.

Active tracking

Strong evidence about the origin of an attack can be gathered by tracing back the attacker to the IT system where the connections or the controlling commands for the attack originate. Common attack approaches often use a so-called command and control infrastructure (C2 or C&C), where specific computers are used to coordinate the attack and collect the stolen data. In order to identify the attacker it is necessary to monitor and gather information about user operations from these specific systems either through hacking them or through international cooperation with the states where the compromised devices are located. The former strategy is known as 'hack back' or 'active defense' and has drawbacks that need to be considered.[9] These disadvantages are for example misinterpretations and wrongful attribution due to insufficient information, the risks of falling for deliberately created 'false flags' and the question whether the attributed system had been used intentionally for the attack or whether it had been exploited.[10] Another approach that enables monitoring an attack but avoids the risks of hack back is to deliberately become one of the exploited systems that the attacker is using – similar to the honeypot approach.

Assembling the puzzle

All these approaches can help a defender to collect data and information about the tactics, the tools and the different steps of an attack in order to compare them to known capacities of threat actors and the sophistication and methods attested in former incidents.[11] It is important to bear in mind that while each of these individual pieces of information can

[7] After incidents, detailed information about the tools of the 'defending' side are rarely revealed. Therefore, it is difficult to point out a real-world example of honeypot usage. Press reports covering the recent attack against the Federal Foreign Office in Germany however stated that the investigating agencies are aware of the incident and are monitoring the attackers' activities which may be an indication that tools like honeypots or beacons had been used. For more details see "Cyber-Espionage Hits Berlin - The Breach from the East," *Der Spiegel* (Spiegel Staff, 2018)

[8] Honeypots can also be installed as a preventive measure but are most effective when tailor-made to a specific attack and its anticipated goals.

[9] See Thomas Reinhold and Matthias Schulze, " Digitale Gegenangriffe – Eine Analyse der technischen und politischen Implikationen von „hack backs"" (Reinhold & Schulze, 2017)

[10] A common and slightly overused example is that of a hospital IT system that may have been hacked itself and used by attacker as hub to indirectly perform another cyberattack. Any offensive countermeasures that disrupt the hospital's services would impair important primary tasks and could result in injuries to human life.

[11] It is important to point out that although still only a limited number of state actors have sufficient offensive cyber capacities, their number is rising. For example, North Korea has developed significant cyber power over the last year with – compared to conventional military armament – few financial resources.

be a lead to the attacker, they can also be manipulated or crafted to leave misleading tracks which could potentially incriminate a third party. A consolidated and coherent analysis needs data collected through a range of various measures. It is certainly possible to conduct such a technical analysis when time is not a problem.[12] While in certain scenarios, such as espionage operations, attribution of an attack might not be time-sensitive, other instances exist where time is a critical factor – for example if a hack back needs to be conducted. Moreover, during military conflict, time might be of the essence but thorough technical attribution takes time and needs to be complemented by an analysis of the geopolitical context in which the attack takes place as well as by intelligence findings.

11.2.2 *Intelligence aspect of attribution*

Obtaining all kinds of intelligence, especially human intelligence and signal intelligence, is crucial to help establish the attribution of cyberattacks. Such intelligence can be gathered through a state's own means or accessed via shared resources by allies. Intelligence can help to attribute one attack or an entire set of attacks in combination with the technical aspects. If for example a technical attribution analysis reveals that certain cyber operations are linked to each other – e. g. because they rely on the same infrastructure – and intelligence can link one of those operations to the perpetrator, an entire campaign of cyberattacks might be unraveled. The indictment filed by the US Special Counsel investigator Robert Mueller for example, in the inquiry into Russian interference in the presidential election, shows that access to email accounts provided the investigators with useful intelligence, enabling them to connect certain dots.[13] Additionally, the public learned that America's National Security Agency is actively tracking various cyber threat actors via signals intelligence tools.[14] In the case of the attack on Sony Pictures Entertainment,[15] it was rumored that American intelligence agencies had access to the network from which the attack originated and therefore were swiftly able to attribute it to North Korea. More information was revealed about Dutch intelligence services which were tracking the Russian hacking group 'Cozy Bear' at least between 2014 and 2015.[16] Hackers from the domestic Dutch intelligence agency AIVD were able to witness and monitor the launch of cyberattacks against the Democratic National Committee[17] because they had access to the network from which this operation was launched. AIVD also had access to security cameras monitoring the offices from which those attacks were conducted, conveniently allowing them to compare the pictures taken with those of known spies. This operation is likely responsible for the strongest proof of a Russian cyber aggression that has ever been obtained and found its way into the

[12] An example is the 2013 Mandiant report "APT1: Exposing One of China's Cyber Espionage Units" which analyzed and presented forensically detailed data and evidence about the Chinese state-driven cyber espionage program about the PLA Unit 61398. (Mandiant Corporation, 2013)

[13] United States of America v. Internet Research Agency LLC et al., Case 1:18-cr-00032-DLF, filed on 16 February 16, 2018 (US Department of Justice, 2018)

[14] Kim Zetter, "Leaked Files Show How the NSA Tracks Other Countries' Hackers," *The Intercept* (Zetter, 2018)

[15] Andrea Peterson, "The Sony Pictures hack explained," *Washington Post* (Peterson, 2021)

[16] Huib Modderkolk, "Dutch Agencies Provide Crucial Intel about Russia's Interference in US Elections," *de Volksrant* (Modderkolk, 2018)

[17] Sven Herpig, "Cyber Operations: Defending Political IT-Infrastructures. A comparative problem analysis supported by the Transatlantic Cyber Forum," (Herpig, 2017)

public sphere. Although crucial to solving the challenge, the intelligence component has been the most underrated aspect in the public debate. The reason for that is the classification of intelligence materials and thus their rare exposure to public scrutiny. After the US presidential elections in 2016, the American intelligence community issued a declassified intelligence report[18] that was supposed to convince the public of Russia's guilt. It however achieved almost the opposite effect because – due to declassification – the public report no longer contained any hard proof of Russian intervention. When asked whether they think Russia attempted to meddle in the 2016 presidential elections, 45% of respondents in the US answered either that they do not know or that it is not true.[19] At the end of the day, it is the state's strategic choice how much it discloses about what it knows and how it obtained its intelligence. Therefore, credible attribution is indeed within the realms of possibility. Whether that proof can be presented to international organizations (e.g. UN, NATO) and/or the public or not is a different story as this would likely mean exposure of the intelligence operation. Revealing such an intelligence operation would decrease the likelihood of it still being effective in the future. If attackers follow the counter-response to their actions closely, they might be able to identify what measures were used to track them down and circumvent/avoid them if possible.

11.2.3 *Geopolitics of attribution*

Geopolitics might only play a minor role in the attribution of cyber operations but this dimension should not be disregarded. While a thorough analysis of the technical aspects and solid intelligence can clearly provide hard facts and concrete evidence when it comes to attribution, a geopolitical assessment can help validate the overall process of attribution. A geopolitical assessment ultimately focuses on the attacker's motivation and hinges on two questions: *cui bono*? (' 'who benefits?') and 'was it a "false flag"[20] operation?'[21] It is rare for an actor to take responsibility[22] for a cyberattack. Even then, the admission has to be vetted and treated with a certain amount of skepticism because it might just be part of a deception strategy. *Cui bono* asks the question of who would directly and most significantly benefit from the attack. Such an analysis can factor in various political aspects, such as ongoing conflicts, current negotiations or recent events. Findings of the technical analysis, such as what documents were stolen and which positions the employees whose computers were breached held in the organization, add value to an assessment. A major reason why Russia has been blamed for so many attacks in recent years is that it stood to gain from all of them, assuming that Russia's main goals are to destabilize Western democracies and project power partly in an endeavor by the Kremlin to divert attention from the country's own domestic problems. The shortcoming of that assumption in terms of attribution is that it is overly broad and therefore involves the risk that Russia is automatically blamed for most cyberattacks. The second aspect of

[18] Office of the Director of National Intelligence, "Background to 'Assessing Russian Activities and Intentions in Recent US Elections': The Analytic Process and Cyber Incident Attribution," (US-DNI, 2017)

[19] IPSOS/REUTERS Poll Data (Ipsos Public Affairs, 2018)

[20] An attack which while disguising the real perpetrator creates the impression that a third party is behind it.

[21] A third aspect could be 'for lulz' (for fun). While this kind of motivation has been in sharp decline in the past few years, groups such as Anonymous and LulzSsec have conducted a number of high-profile hacking operations with the apparent goal of ridiculing the victim.

[22] Noah Shachtman, "Kremlin Kids: We Launched the Estonian Cyber War," *Wired* (Shachtman, 2009)

a geopolitical assessment, false flag operations,[23] is straightforward because it asks a similar question: who benefits from the cyberattack in a case where another and/or the most obvious actor identified by a *cui bono* assessment will be blamed for the attack? Many of the indicators examined in a technical analysis, such as timestamps, language configurations, comments or hidden pictures in the code, can be easily manipulated to point in a certain direction. Using a geopolitical cover at the same time makes a false flag operation even more effective. A notable example of this was the alleged Russian cyber operation 'Olympic Destroyer' which not only relied on borrowing technical elements from previous North Korean cyber operations but targeted the Winter Olympics in South Korea[24] at a crucial moment in the North Korean-South Korean and American diplomatic relationship. If Russia was indeed behind it, this false flag operation was definitely smart. If North Korea is blamed for the attack, the relationship between the two Korea's would further deteriorate, forcing the United States to devote more of its attention to that part of the world (instead of towards Russia). If Russia is blamed for the attack, it can further its agenda of power projection and at the same time undermine public confidence in attribution in democratic countries.[25] False flag operations add an additional layer of complexity to an already complex phenomenon.

11.3 Why It Is Problematic to Point to Russia

Russia has repeatedly been blamed for cyberattacks in the past decade. Every other operation is currently linked to Russia by politicians, the media or IT security companies.[26] Some of the attribution might be correct, some might be wrong. The challenge here is not attribution but *credible attribution*. Credible attribution does not only mean getting the technical, geopolitical and intelligence aspects of attribution right, it also means convincing the target audience.

Whereas attributing the source of an armed attack is possible for missiles or conventional military forces, such attribution is arguably nearly impossible or considered impracticable[27] in the case of attacks carried out via cyberspace as described earlier. Cyberspace offers perfect conditions for attackers to obfuscate their tracks and deceive the defenders and forensics. Attackers could use uninvolved third party IT infrastructure – or could fly to a different country with a 'burner laptop'[28] – to conduct an attack. A targeted victim can only immediately identify the last element of the chain of computers used

[23] NATO Cooperative Cyber Defense Center of Excellence (CCDCOE), "Mitigating Risks arising from False-Flag and No-Flag Cyber Attacks" (CCDCOE, 2018)

[24] Andy Greenberg, "Russian Hacker False Flags Work - Even After They're Exposed," *Wired* (Greenberg, 2018)

[25] Levi Maxey, "False Flags in Cyberspace: Targeting Public Opinion and Political Will," *The Cipher Brief* (Maxey, 2018)

[26] See for example the FireEye report from 2014 on the APT28 group; "APT28: A WINDOW INTO RUSSIA'S CYBER ESPIONAGE OPERATIONS?," (FireEye, 2014) as well as the CrowdStrike report from 2016 "Who Is COZY BEAR?," (CrowdStrike, 2016)

[27] See the conclusions of the 2016 UNIDIR report of the International Security Cyber Issues Workshop Series (UNIDIR, 2016)

[28] A device which is only used for a particular attack and then trashed to hinder attribution. Derived from the concept of a 'burner phone'.

in the attack but not the origin behind it.[29] Strong empirical attribution would need to identify every device in the attack chain, and gather and analyze available traces to forensically link them to the real origin of the attack. This is a complex task[30] which is challenging even under optimal conditions where every IT system within the described chain contains traces of the attacker and the victim is able to gather these data via international cooperation.[31] Such a task will not work in specific conflict situations where an immediate response is necessary and 'conclusions shortcuts' are dangerous because the ambiguity and incompleteness of the information about a cyberattack raises the risks of misunderstandings, miscalculations, misinterpretations and wrong responses, especially when other means of crisis communication or confidence-building measures between the adversaries are missing. Besides such 'hard facts', prior events have shown that attribution is still ultimately a political decision based on information collated by intelligence and security agencies or influenced by foreign policy interests and considerations[32]. There are only very few instances in which states based a public response, e.g. sanctions, on the findings of an attribution assessment. One of them was the US response to Russia's **alleged** meddling in the 2016 presidential election campaign.[33]

Additionally, states that are blamed for an attack often distance themselves from the hacking group that conducted the operation and deny any official involvement or control of the group. Even though the UN GGE decided to hold states accountable for cyber operations conducted from within their territory,[34] pledging to help the investigation with any means possible will take some pressure off a state that finds itself under suspicion. Plus, linking a cyberattack to a hacker group is one thing, linking that hacker group or a specific incident to a state and especially to a particular governmental or military order as is required by the UN Charter is quite another. Even if due diligence is a commonly accepted principle in cyberspace[35] it is not enforced in the current public debates on potential cyberattacks from Russia. In fact, prior to the establishment of a military cyber unit in 2017, the Federal Security Service (FSB) was responsible for overseeing Russia's cyber capabilities.

From the perspective of the international community and as described earlier in respect to international law, attribution and accusations in specific conflicts need to be based on a credible, evidence-based argumentation that has to be made by the affected state. But so far no case exists where such evidence that points inexorably to Russia had

[29] An attack might have several 'origins', which are intermediate systems exploited by the attacker to make an uninvolved third party look like the adversary. The 'real origin' of an attack is the point where the attack was started by the aggressor.

[30] Two case studies that show the complexity of this task, the different sources that have to be taken into account, the technical difficulties and challenges of tying this information together are the final report of Ralph Langner on Stuxnet, "To Kill a Centrifuge", (Langner, 2013), as well as the 2013 Mandiant report "APT1 - Exposing One of China's Cyber Espionage Units," (Mandiant Corporation, 2013)

[31] A good example of the complexity of this task is given in Ralph Langner's analysis of the Stuxnet incident. See Ralph Langner, "To Kill a Centrifuge", (Langner, 2013)

[32] For instance the hacking attacks against German governmental and parliamentary IT systems from 2015 and 2018 yielded no official reaction against the suggested attackers, whereas a hacking attack against the US-based company Sony Pictures Entertainment from 2014 almost immediately (in terms of days) resulted in US sanctions against North Korea.

[33] David E. Sanger, "Obama Strikes Back at Russia for Election Hacking," *New York Times,* (Sanger, 2016)

[34] NATO Cooperative Cyber Defence Centre of Excellence (CCDCOE), "2015 UN GGE Report: Major Players Recommending Norms of Behavior, Highlighting Aspects of International Law," (Rõigas & Minárik, 2015)

[35] See Annegret Bendiek, "Sorgfaltsverantwortung im Cyberraum - Leitlinien für eine deutsche Cyber-Außen- und Sicherheitspolitik, " (Bendiek, 2016)

been made public. Two instances which provided the most public information about a state-backed attribution pointing towards Russia are the US Director of National Intelligence's report[36] and the Dutch domestic intelligence AIVD findings.[37] Enabling states to conduct more severe responses to cyberattacks would require public and international attribution. This in turn leads to a Catch-22 situation in intelligence sharing. Pointing the finger at the usual suspect without credible attribution when a cyberattack occurs not only further emboldens Russia in its projection of power (while it continues to deny responsibility) but fails to sufficiently convince the public or the international community.

This pattern of accusation and denial underlines again the necessity for binding international rules of state behavior in cyberspace that include a commitment to the validity of due diligence principles in this domain. This would provide a strong basis for enforceable regimes of international law in cyberspace.

[36]Office of the Director of National Intelligence, "Background to 'Assessing Russian Activities and Intentions in Recent US Elections': The Analytic Process and Cyber Incident Attribution," (US-DNI, 2017)

[37]Huib Modderkolk, "Dutch agencies provide crucial intel about Russia's interference in US elections" (Modderkolk, 2018)

12

WANNACRY ABOUT THE TRAGEDY OF THE COMMONS? GAME-THEORY AND THE FAILURE OF GLOBAL VULNERABILITY DISCLOSURE

Abstract Vulnerabilities in Soft- and Hardware have global implications in an interconnected world, since they affect every user who uses an affected system. Since cyberattacks relying on software vulnerabilities produce significant costs for national economies and societies, finding and closing these vulnerabilities is in the rational interest of many countries. Coordinating vulnerability disclosure and timely patching on global scale thus would be a common interest shared by all states. However, states in particular withhold software vulnerabilities for the purpose of foreign espionage, surveillance and law enforcement. Thus, common and particular interests collide, resulting in what game theory calls a tragedy of the commons. Global cyberspace becomes more insecure as more and more states withhold critical software vulnerabilities. In game theoretic terms, rational-actions on a local level produce irrational effects on a global scale, representing a prisoners dilemma. The paper uses game theory to develop a set of international best practices to escape the prisoners dilemma of software vulnerabilities. The questions thus becomes, *what the smallest common denominator of such a global vulnerability disclosure regime could be and under what conditions can such an agreement could reached.* The case for developing these proposition is the EternalBlue incident of 2017, a software vulnerability that was hold back by an intelligence service and whose unintended disclosure resulted in several destructive malware campaigns with economic damage on a global scale.

Original Publication *Schulze, M., & Reinhold, T. (2018). Wannacry About the Tragedy of the Commons? Game-Theory and the Failure of Global Vulnerability Disclosure.* European Conference on Information Warfare and Security, ECCWS, 2018-June, *454–463. https://www.proquest.com/openview/f6ccddd62973bd8997c3fcd40951f4f1/1?cbl=396497&pq-origsite=gscholar&parentSessionId=7jm9tc94UMKaTk9pATjzd%2BhJYdl8V55qGHqrUpnUM8%3D*

12.1 Introduction

Finding and patching software vulnerabilities becomes is central issue to prevent the exploitation by hackers. A software vulnerability that is unknown to the vendor is usually called a zero- or 0-day vulnerability (Ablon & Bogart, 2017). To find these, software developers rely on responsible disclosure programs, where researchers and white-hat hackers can disclose a vulnerability they find to the vendor (Shepherd, 2003). The vendor usually then has up to six months to develop a patch that fixes the disclosed vulnerability before it is made public. This vulnerability-discovery, -disclosure, -patching and patch-

installing cycle is one of the core defensive cybersecurity measures (Frei, 2013). In an interconnected world, one unpatched vulnerability, like in commonly used Windows systems, potentially affects billions of users worldwide.

That one vulnerability could be theoretically used to access the majority of interconnected computer systems worldwide drives the cost-benefit calculus of cyberattackers. Like researchers, hackers engage in the costly and time-consuming process to hunt for vulnerabilities in software that can turned into an exploit that grants remote access to computers or networks. Alternatively, hackers can buy off-the-shelf 0-day vulnerabilities or ready-made 0-day exploits from vulnerability brokers on international gray markets (Ablon & Bogart, 2017). Thus, in order not to waste financial resources and developing time, cyberattackers like intelligence agencies have an inherent interest in withholding or stockpiling vulnerabilities for offensive cyber-operations (Healey, 2017). Stockpiling means that the knowledge about this vulnerability is not disclosed to the vendor, thus no patch exists. Cyber-defenders want to close every possible crack in their systems to prevent intrusion. This offense-defense dilemma becomes more problematic as more actors enter the cyberwarfare arms race. The more states engage in cyber-operations, the more vulnerabilities will be stockpiled, thus remain open for exploitation. By withholding vulnerabilities, states risk that another actor finds the very same vulnerability and exploits it. Thus, states engaging in cyber-operations have conflicting interests: stockpiling vulnerabilities for a national advantage or disclosing vulnerabilities to increase cybersecurity for everyone.

The aim of this paper is to analyze the conflicting offensive and defensive interests in greater detail. The research question is how states can overcome their conflict interests in order to achieve cooperation in discovering and disclosing software vulnerabilities to increase global cybersecurity. Since national interests and questions of cooperation and conflict lie at the core of international relations (IR), we utilize IR theories and game-theory to sketch out the vulnerability dynamics. To indicate the global security risks of software vulnerabilities, we rely on the WannaCry malware incident from 2017 as a case study. Finally, we offer some preliminary strategies to achieve cooperation in disclosing vulnerabilities on a global scale.

12.2 COOPERATION UNDER ANARCHY

How to resolve conflict of interests, in order to achieve cooperation among two entities under the conditions of anarchy, has been at the core of IR theory. The central challenge for cooperation is the anarchic structure of inter-state relations, understood broadly as a lack of a world government that can enforce global laws. Actors under anarchy try to realize their pre-defined interests like survival, sovereignty or wealth. Since they are lacking an overarching governance structure that coordinates their behavior, competing interests, for example in acquiring tangible scarce resources (like oil or rare metals) or intangible goods such as security, often result in conflicts of interest (Axelrod & Keohane, 1985). For example, the independent self-interest of neighboring maritime actors in fishing in common waters might deplete fish stocks. Spoiling scarce resources through collective action became to be known as the tragedy of the commons (Stein, 1982). The individual interest for fishing ranks higher as the collective interest of maintaining common fish stocks in the preference order of actors. In security politics, state A's

interest in achieving security often results in unilateral armament. Since most weapons can be used offensively, a neighboring state B cannot be sure about the intentions of the other and thus might perceive a decrease of its own security, which is against its interest. Fear, mistrust and uncertainty about intentions might lead to a vicious cycle in which B in turn starts to arm itself, increasing perceived insecurity for state A. This has been called a security dilemma (Tang, 2009). The same logic can be applied to the practice of states hoarding zero-day exploits for espionage or law-enforcement purposes instead of disclosing them with the manufacturer. In order to do so, the next chapter will introduce the structural logic of such a dilemma.

12.2.1 *Prisoners Dilemma and Zero Days*

In game theory, the Prisoner's dilemma has been used to describe situations in which the realization of individual self-interest produces suboptimal collective outcomes. S.M. Amadae describes the standard textbook narrative of the dilemma:

Ÿou and your co-conspirator have been captured by the authorities. You are separated and each given the choice between confessing and remaining silent. One of four possible outcomes will occur: If you confess while your partner remains silent you go free. If you both remain silent, you each receive one year in prison. If you both confess, you each receive a five-year sentence. If you remain silent while your partner confesses, you face a ten-year sentence while your partner goes free. What do you do?¨ (Amadae, 2016)

Both conspirators agree to remain silent before going to prison, but once incarcerated cannot communicate with each other. Both players in the game can either confess (defect) or remain silent (cooperate) resulting in four possible outcomes. The Prisoner's Dilemma is typically depicted like the following figure:

		Player B	
		Cooperate	**D**efect
Player A	**C**ooperate	2 / 2	0 / 3
	Defect	3 / 0	1 / 1

Table 12.1: Prisoner's Dilemma

The rows represent the preferences of player B, while the columns indicate A's preferences. The numbers represent the order of preferences. The higher the number, the higher the payoff.[1] The preference order for both players is DC>CC>DD>CD. The dilemma lies in the preference structure of the game, which produces an incentive to defect no matter what the other one does, resulting in both ending up worse (1/1) than if they had cooperated (2/2). The key issue about this game is uncertainty because both players cannot be sure whether the other remains complicit or defects, which is why it has been used to depict arms races.

The same logic applies to the strategic rationality behind hoarding zero day exploits. In these instances, states could either cooperate (2/2) by responsibly disclosing vul-

[1]The first number represents player A and the second player B.

nerabilities to vendors. If the vendor fixes the vulnerability and rolls out the update, an increased level of cybersecurity is provided. The cybersecurity created by software updates has the character of a collective good or a commons from which everyone, using the same hard- or software, can benefit as a free rider (Stein, 1982). Alternatively, an actor chooses to defect from this practice and withholds the vulnerability (3/0). The rationality behind this practice is that if the antagonist has not discovered the same vulnerability and fixed it, it allows breaking into its computer systems or networks without getting detected. Thus, withholding vulnerabilities represents a clear advantage for the cyber-offense, which explains why both actors prefer their own defection over mutual cooperation under uncertainty. Because both actors have an incentive to defect, they arrive at the suboptimal outcome (2/2) that both defect and withhold a certain amount of zero-day vulnerabilities for their cyber-arsenals. Thus, a certain degree of vulnerabilities remains unfixed and could be potentially exploited by every actor, state and non-state, who identifies the same vulnerability. In sum, withholding vulnerabilities for national security reasons implies a lower degree of common cybersecurity.

The next chapters illuminate different theoretical approaches to escape the vulnerability dilemma and test these propositions towards the particularities of cyberspace.

12.2.2 *Escaping the Dilemma*

The Prisoner's dilemma can be solved either with game-theoretic mechanisms or by alternative theories of (bounded-) rationality (Kahneman, 2003) or normative theories (Finnemore, 2017). Because of the limited scope of this paper, we will focus on game-theoretical mechanisms for escaping the prisoners dilemma. Axelrod and Keohane describe three ways: changing the payoff structure of the game, the shadow of the future, and reciprocity in regime structures (Axelrod & Keohane, 1985).

The payoff structure determines cooperation. The more intense the conflict of interest, the less likely cooperation becomes. This often depends on factors outside the control of actors, like for example market prizes for 0-day exploits or contextual factors like economic recessions or power structures. Jervis argues that changes in the payoff structure are possible with policies that aim at either ïncreasing the mutual gains of cooperation (CC)¨ (Jervis, 1978) and/or decreasing the costs of a player being exploited. Alternative ways to cooperation would be, decreasing the gains of defection, increasing the costs of mutual noncooperation (DD), or lastly, increasing the expectation that the other will cooperate (Jervis, 1978). Higher costs of being exploited, like a loss of sovereignty, increase the security dilemma, whereas tolerable costs make security easier to attain. Exploitation costs can be reduced by building resilience that compensates for defection or by increasing defensive mechanisms, that make offensive moves more costly (deterrence by denial).

A more challenging approach is to increase the objective gains of mutual cooperation and the subjective perception that cooperation is beneficial. If actors adopt a zero-sum approach to politics by ignoring mutual interdependence, cooperation becomes less likely (Amadae, 2016). Changing perceptions within state A requires plausible signaling that actor B is indeed planning to cooperate, as well as a communication of the expected gains and costs in case defection occurs (Jervis, 1978). In international security, good-

faith diplomacy and confidence building measures serve this purpose. The ultimate form of good faith would include unilateral disarmament. Another way could be driving up the costs for mutual defection that it becomes an unfeasible scenario. The theory of mutually assured destruction (MAD) in nuclear security strategy had this purpose by threatening a full scale reaction towards a nuclear first strike. This altered the general payoff structure and turned it into a chicken game (Amadae, 2016). In a chicken game, both players drive towards a cliff. Both players win of one of them unilaterally yields. In other words, being exploited (CD) is not as bad as both players driving down the cliff (DD).

Another way is changing the game structure into an iterated game, that allows reciprocal or tit-for-tat behavior: each actor cooperates as long as the other does so (Grieco, 1988). In theory, this enables cooperation by introducing the shadow of the future. The fear of future retaliation for today's defection drives much of security policy, especially in the area of deterrence by punishment. If future payoffs are more valued than the incentive to defect today, cooperation becomes more likely. In turn, if both players believe cooperation is worse than defection, the game results in deadlock. The key challenges in repetitious games are the ability to detect and attribute defection and to impose sanctions aimed to prevent future defections. Detection requires reliable information about the other players incentives, motives and actions (Axelrod & Keohane, 1985). If players cannot identify defectors, are unable to focus retaliatory measures or have no long term incentives to retaliate, effective reciprocity is stifled by what is called a s̈anctioning problem¨ (Axelrod & Keohane, 1985). In the Prisoner's dilemma, the lack of information about the players motives and transparency of actions is structurally forced upon the players. Thus, sanctioning problems are particularly complicated in areas involving secrecy and clandestine behavior, such as the intelligence and espionage. It gets even more complicated in games involving more than two players.

Sanctioning problems have been historically resolved by creating international regimes that manage participants expectations, standardize behavior, thus provide information about compliance and assign the r̈esponsibility for applying sanctions¨ (Axelrod & Keohane, 1985). Regimes are typically defied as s̈ets of implicit or explicit principles, norms, rules, and decision-making procedures around which actors' expectations converge in a given area of international relations.¨ (Krasner, 1983) The major function of regimes is to transform independent, self-interested decision-making into joint decision-making (Stein, 1982). Regimes must define what actions specifically constitute cooperation and defection and impose compliance mechanisms (Stein, 1982). Grieco argues that ïf the costs of verifying one another's compliance, and of sanctioning cheaters, are low compared to the benefits of joint actions, then cooperation is more likely (Grieco, 1988).

The aim of the paper is not to design a full-fledged vulnerability regime including questions of regime design (Young, 1980), efficiency or dynamics (Young, 1982), but rather to sketch out minimal cooperation preferences that could be developed further. For that, unique aspects of cyberspace that might influence preferences and thus possibilities must be discussed.

12.3 Cooperating on Vulnerability Disclosure

The global nature of an interconnected cyberspace implies a necessity for actors to cooperate because unilateral actions to increase security rarely work in this domain. State actors have not yet realized the potential gain for cooperating in vulnerability disclosure on a global scale. To explain that, one needs to keep in mind certain particularities of cyberspace that reduce the effectiveness of traditional approaches of international regimes like arms control treaties.

12.3.1 *Particularities of cyberspace and digital goods*

The digital sphere exhibits many of the features that make cooperation even more complicated. Any approach that proposes a change of the costs of games needs a clear estimation of potential costs. This includes the vulnerability of one's own infrastructure and that of adversaries, as well an assessment of the potential for destruction (in terms of range and impact) of a specific vulnerability exploit. Achieving situational awareness is difficult because it requires knowledge of all systems – civil, military and government – in a given state, introducing a full range of privacy and feasibility issues. The general issue with the digital world is that is mostly run and occupied by private entities with diverging interests. The same is true for software vulnerabilities which concern industry, researcher, criminals and states alike, making cooperative behavior more complicated (Arimatsu, 2012).

Another problem concerns measurement of actions and effects. Tangible goods like fish-stocks are immediately measurable. Like dual-use software in general, IT-vulnerabilities are intangible goods that have no essential characteristics that can be quantified or qualified in a coherent way. Software vulnerabilities the character of knowledge. Regulating or prohibiting the proliferation of know-how is inherently challenging, if not impossible while cherishing norms such as freedom of speech. That is why, to this date, attempts to define and regulate c̈yberweaponsḧave resulted in no international consensus (Arimatsu, 2012). One particular feature of software vulnerabilities is their üse and loosec̈haracter. The disclosure and patching of an exploit renders it useless, which implies a limited half-life (Ablon & Bogart, 2017). This characteristic and a general belief in the superiority of the cyber-offense vis-à-vis the defense, drives actors towards using exploits before they expire (Singer & Friedman, 2014). Under such conditions, cooperation becomes harder (Müller, 1982).

Another specific feature of the cyberspace is, that it offers many ways for actors to hide its tracks and therefore be able to deny responsibility and avoid sanctions when using exploits for espionage etc.. The attribution problem affects the measures of verification and compliance control to agreements because actors can effectively use proxies that are hard to retrace (Clark & Landau, 2011). Even cooperation can be undermined by keeping back specific exploits because the missing mutual verification measures in cyberspace make it impossible to gain assurance over the reliability of agreements.

Thus, achieving cooperation in cyberspace is particularly hard but important nonetheless. The next chapter introduces the WannaCry case to sketch out the global implications of vulnerability stockpiling for national security.

12.3.2 *The WannaCry and NotPetya malware campaigns and its EternalBlue roots*

"WannaCry" was a ransomware campaign that was launched in may 2017 and infected approximately 230.000 IT systems in over 150 countries world wide (Oberhaus, 2018). The malware could spread and infect the computers very efficiently by exploiting a critical security hole in older and unpatched Microsoft Windows version. In June 2017, another malware campaign dubbed "NotPetya" infected computers in Ukraine, Russia, the USA and some European countries by using the same security hole. This malware also encrypted the files but, in contrast to WannaCry, it deliberately tried to sabotage the computers. The exploit that both malware campaigns used is known as EternalBlue (CVE-2017-0144) and was stolen from the National Security Agency (NSA) by a hacking group named "Shadow Brokers" during the end of 2016 till the first weeks of 2017. The exploit was leaked by the hacker group on April 14, 2017.

According to Microsoft, the NSA informed the company about the exploit after they learned about its theft. Microsoft published a security patch for the officially supported Windows versions in March 2017 and - after the WannaCry impact - even re-established old, formerly deactivated patching channels to supply emergency security fixes for the old and unsupported but yet still popular Windows versions like XP. Even with these patches, both malware campaigns produced a lot of damage.

One of the worst impacts of WannaCry were the infections of hospitals of the National Health Service in England and Scotland, Nissan Motor Manufacturing UK, the French car company Renault, Spain's Telefónica, the German Deutsche Bahn and the US company FedEx. The Trend Micro's security and threats report estimates the loss of up to 4 billion USD (Trendmicro, 2017) and the official case analysis of the UK National Health Service counts "*thousands of appointments and operations [that] were cancelled and in five areas patients had to travel further to accident and emergency departments*" (Sir Amyas Morse KCB, 2018).

The NotPetya attacks affected several Ukrainian ministries, banks and metro systems but also international companies like the Danish Maersk Line, Russian oil company Rosneft, the German DHL, FedEx and the hospital operator Heritage Valley Health System. FedEx estimated a loss of $300 million USD due to its affected daughter company TNT Express (Schlangenstein, 2017). The Danish Maersk Line, that, according to their own description, handles around 20% of the worldwide container shipping (Chirgwin, 2018), estimates between 200 to 300 million USD financial damage (Mathews, 2017). NotPetya's financial impact is estimated at $892 million dollars (O'Connor, 2017). These examples show that the leak of the EternalBlue exploit affected the worldwide economy as well as the US economy itself, even despite the intervention of the NSA and the patch rollout of Microsoft.

The EternalBlue case shows distinctively the competing interests of states regarding software vulnerabilities. Before it got stolen, NSA cyber-operators allegedly used this

vulnerability for more than five years to gather foreign intelligence. According to anonymous NSA employees, ẗhe intelligence haul was unrealänd ït was like fishing with dynamite, indication of great utility (Nakashima & Timberg, 2017). Since it affected most Windows versions, it was like a general key to the digital world. EternalBlue decreased the security of the United States because its proliferation aided two of its adversaries – North Korea and Russia – by increasing their know how for their cyber-operations (Heller, 2018; Palmer, 2017). Not just nation-states utilized EternalBlue for cyber-operations, but also multiple criminal groups repurposed this malware. The banking trojan Retefe (PCriskcom, 2017; Threatpost, 2017), the crypto-currency miners Adylkuzz, WannaMine, and other malware such as Uiwix, EternalRocks or the ransomware Bad Rabbit also utilized the same vulnerability (ENISA, 2017). Researchers from Proofpoint also found a crypto-currency-mining Botnet based on EternalBLue called Smominru, consisting of 526,000 Windows computers (Khandelwal, 2018). Intangible and indirect costs like psychological effects like a loss of trust in digital infrastructures are hard to calculate, but should also be considered.

12.4 SUGGESTIONS TO FOSTER THE COOPERATION IN CYBERSPACE

What can be learned from the game theoretical approaches with regard to the specific constraints of cyberspace? An easy-to-reach incentive for cooperation can be mutual agreements about threat and vulnerability sharing between partners, which resembles a reiterated or tit-for-tat cooperation. Such a process exists in the form of computer emergency response teams (CERT) and is getting extended and formalized in the last years by the European Union Agency for Network and Information Security (ENISA). Currently, these platforms are used to collect and inform partners about specific threats and provide a shared level of situational awareness, but the concept could be extended to share exploit and vulnerability information too, if partners could agree on detailed rules what kind of exploit information - in terms of the discussed specific threat categories - will be shared. Such an approach requires the standardization of vulnerability disclosure mechanisms like the Common Vulnerability Exposure (CVE 20918). Such platforms hugely benefit those participating in them, and thus could also be utilized to raise the costs of non-cooperation by the exclusiveness of time-critical threat warnings, where non-partners will get informed only with the official patch release and the vulnerability warning of the manufacturers.

Alternatively, one could think of an internationally accepted information platform where partners do not share the exploits and vulnerabilities themselves, but the mere existence and/or the purchase and trade of these kinds of information. A similar approach had been done by the extension of the Wassenaar-Agreements in 2013 by including “intrusion software” in the list of regulated goods (Granick & Fidler, 2014), although this only affects specific software or hardware but not the vulnerability knowledge behind.

A different approach is to lower the benefits of non-cooperation by shortening the life cycle and thus the utility of vulnerabilities. This could be done via binding legal definition and procedures how long an actor could keep the knowledge of vulnerabilities until it needs to get disclosed - independently how the knowledge was gained in the first place (own research and analysis or bought from a vendor). This approach, that is currently internationally discussed as “vulnerability equity process” (VEP) (US White

House, 2017) accepts that some actors have justified needs for vulnerabilities but will on the other hand dampen the threats of this situation. One challenge of this measure lies in the necessity of attendant verification regimes to ensure the compliance to these agreements.

Another way to reduce the utility of vulnerabilities would be the establishment of an international process to re-finance the public and scientific research on IT security. Simple suggestions would include mandatory and internationally standardized bug-bounty programs for all software vendors, which can be economically feasible (Finifter et al., 2013). The general critique of current bug bounties is that the payoff for ethical hackers is comparatively small, compared to the benefits of selling an exploit on the black market. Apple for example offers a maximum of $200.000 for a secure boot exploit, whereas vulnerability brokers such as Zerodium pay up to $1,5 million (C. Miller, 2017).

More radical suggestions include to change the pay-off structure for vulnerability brokers and cyber-criminals by drying out the vulnerability black market. A study by NSS Labs suggests, that for large IT companies the hypothetical costs of purchasing all available vulnerabilities from the market would be significantly lower compared to expected losses as the result of cybercrime utilizing these vulnerabilities (Frei, 2013). They propose an International Vulnerability Purchase Program in which states and corporations would share the financial burden in buying all available vulnerabilities from the market and encourage ethical hackers to disclose them for more competitive rewards than typical bug bounty programs.

Arms control regimes for cyberweapons are typically seen as ineffective because of verification issues, the intangible dual-use nature of digital technologies and the multitude of (non-state) actors (Arimatsu, 2012). Instead of relying on a militarized or securitized logic of cyber-arms control, a way out could be to find solutions outside the realm of national security analogies and concepts. One way could be to draft a vulnerability regime similar to emissions trading regimes (Keohane & Victor, 2011). Like Carbon Dioxide (CO2), vulnerabilities in software could be seen as emissions or byproducts of industrial production that do harm their respective ecosystems by creating negative externalities. The essence of CO2 trading schemes, is that market mechanisms change the payoff structure to create incentives for state *and* non-state actors to cooperate tor reduce emissions. To achieve this, the international community defines a cap, the total amount of tolerable CO2 emissions. Industries can buy a limited amount of certificates that allows them a certain degree of pollution. If they produce less CO2, they can trade unneeded certificates and thus generate profit. If they pollute more, they need to buy additional certificates. This scheme has the advantage that it scales with industry size: large polluters need more certificates, whereas small and mid-size companies need less. A similar mechanism could be introduced in the software market where the large players, with highly complex products with millions of lines of codes that are used in a wide array of services and critical infrastructures can buy a certificate that guarantees them a right to have X-amount of vulnerabilities in their software. For this amount, the company is not liable in case a major cyber-incident occurs. If more than X vulnerabilities are found by researchers, the company must buy the right to have more vulnerabilities in their software. In theory, this increases the incentive of manufacturers to invest more in software quality without harming the competitiveness of small IT companies.

A minimum rule that could be sketched out even with competing interests is that if a nation's cyber-stockpile gets stolen, it must inform the software vendor. This has been the case with the ShadowBroker incident. Similar reporting requirements exist in the nuclear realm with the International Nuclear and Radiological Event Scale in case of a major radiation leak that affects countries or after a major virological outbreak within the global health regime of the World Health Organization. These regimes require that states notify the international community and affected neighboring states, that they, can implement protective measures like screening at airports or provision of vaccines.

Of course there are other ways to raise the costs for non-cooperation, that rely on different sets of theories which could not be tackled here, such as psychological or organizational research (Jervis, 1978).

12.5 CONCLUSION

This paper has shown that the tragedy of the commons applies also to cyberspace when seemingly rational actions are structured in a way that produces suboptimal outcomes to the common good. The decrease of the common good cybersecurity by hoarding 0-day exploits is an example. We have introduced a first set of game-theoretic mechanisms that aim at altering the pay-off structure of vulnerability hoarding, i.e. making cooperation more beneficial, or imposing costs on non-cooperation. Any political measure to regulate cyber-offensive actions should consider the cost benefit calculus and should not impose additional costs on the defensive side. Of course, rational choice theories rely on simplistic models of strategic rationality that might not completely apply to the real world. Theory testing is a necessary next step. Future research should also utilize additional theories of cooperation, from organizational studies, social psychology and norm-research to sketch out additional ways to achieve cooperation in vulnerability sharing. Additionally, the suggestions we made need to be further refined. The drafting of a vulnerability regime is a complex endeavor with many potential pitfalls. Before attempting to do so, future research should answer questions of better verification and compliance, as well as questions of regime design and efficiency to avoid paper tigers that have no real world impact or that gets outrun by the high-speed of digital change.

13

THE DIGITAL DIVIDE IN STATE VULNERABILITY TO SUBMARINE COMMUNICATIONS CABLE FAILURE

Abstract The backbone network of submarine communication cables (SCC) carries 98% of international internet traffic. Coastal and island states strongly depend on this physical internet infrastructure to provide internet connectivity. Although about 100 SCC breakdowns of human or natural origin occur at yearly average, a literature review reveals that there is no global comparison that assesses individual state vulnerability to SCC failure in global comparison. In this article, the global SCC network is modeled based on publicly available data. Besides the analysis of the global network properties, a focus is put on remaining bandwidth capacities in three different failure scenario simulations of SCC breakdowns. As a result, this study identifies 15 highly vulnerable states and overseas territories, and another 28 territories that are classified as partially vulnerable to SCC failures. Since economic market decisions shape the structure of the SCC network, an uneven distribution of redundancies and the resulting vulnerability of disadvantaged economies can be confirmed. Therefore, the study's findings may contribute to a better assessment of the necessity of preventive protection measures of critical telecommunication infrastructures in states and territories characterized by high and medium vulnerability.

Original Publication *Franken, J., Reinhold, T., Reichert, L., & Reuter, C. (2022). The Digital Divide in State Vulnerability to Submarine Communications Cable Failure.* International Journal of Critical Infrastructure Protection, 38. *https://doi.org/10.1016/j.ijcip.2022.100522*

13.1 Introduction

With over four billion users, the internet is the dominant medium of communication of present times (ITU Publications, 2012). Although no general and uniform definition of critical infrastructure on the international level has yet emerged (Newbill, 2019), the communication sector is typically part of the core classifications in most countries and international bodies (Hollick & Katzenbeisser, 2019). The United Nations Office for Disaster Risk Reduction (UNDRR) defines critical infrastructures as "*[t]he physical structures, facilities, networks and other assets which provide services that are essential to the social and economic functioning of a community or society*" (UN, 2016). Internet is ubiquitous, at least in most parts of the world, and modern societies and economies are highly dependent on its provision. Therefore, the physical internet providing infrastructures can be considered critical infrastructures (CCDCOE, 2020; DeNardis, 2012).

The internet is based on a multitude of different physical transmission structures that are essential for its operation, the most important being land-based fiber optic communication cables (LCC) and submarine fiber optic communication cables (SCC). For the transmission of global data traffic, the latter is by far most important: More than 98% of international online communication is handled via fiber optic cables laid in the world's oceans (CCDCOE, 2020; Winseck, 2017). Therefore, the global backbone network of SCCs is indispensable for the worldwide operation of online data exchange (Bischof et al., 2018). Currently, over 80% of the 1.3 million kilometers of active submarine fiber optic cables are located in inaccessible deep sea below 1500 m depth (SubTelForum, 2020b), which makes it impossible for authorities or private companies to ensure continuous surveillance and physical protection of it (Carter et al., 2009). Hence, fiber optic cables are regularly exposed to factors that potentially impair their function. According to Mauldin (Mauldin, 2017), most of the incidents originate from unintended human activity at sea, such as fishing (38%) and drag anchoring (25%), followed by environmental hazards (14%) like seaquakes or underwater currents. Considering over 100 failures of SCCs on yearly average (Mauldin, 2017), it appears obvious that the functionality of the global internet cannot be taken for granted. Although many cable failures can be compensated by longer and slower alternative routes, these are not available in all geographical regions (Xie et al., 2019). The alternative technology of satellite-based internet, which can, in theory, be accessed worldwide, is far from being able to transmit the necessary amount of data to compensate for an SCC (Winseck, 2017). Low earth orbit technologies like SpaceX's Starlink or OneWeb do not yet provide the bandwidth currently needed by whole societies, as they are still in trial phase (S. Chen et al., 2020).

The following example of a cable failure that led to the complete loss of broadband connectivity for an entire territory illustrates the consequences of internet outages: On the archipelago of the Northern Marianas, the only available submarine cable ruptured in 2015 due to an underwater currents, completely cutting off the island from broadband traffic for several days (UN-ESCAP, 2018). Cascading effects caused internet, telephone communication, and air traffic to collapse, along with disruptions in the health, tourism, and education sectors. The U.S. overseas territory with 50,000 inhabitants suffered damage amounting to 21 million USD. Small island developing states (SIDS) hardly offer any possibility to operate a cable economically due to their characteristics, such as their remote location, small number of citizens, and below-average GDP, resulting in lower internet usage (Sutherland, 2009). Consequently, if any, internet connectivity is usually available only via one or two submarine cables (ITU Publications, 2019a). Here, the state's dependency on the functioning of an SCC is apparent. Nevertheless, cable ruptures in the past also triggered consequences for countries with multiple alternative cables (Aceto et al., 2018; Kitamura et al., 2007). In order to generate a broader picture and not reduce the consequences of a cable break to the, according to literature, most vulnerable group of SIDS alone (Bueger & Liebetrau, 2021; Sutherland, 2009), a global focus is chosen for this paper.

The research question underlying the work will therefore be:

Which states and overseas territories are vulnerable regarding the loss of functionality of adjacent submarine communication cables in global comparison in mid-2020?

To approach this question, we divided this paper into six sections. After the introduction (Sec. 1), we provide an overview of the related work (Sec. 2). Subsequently, we proceed

with the method section, where core definitions, network analysis tools, and the data compilation are presented (Sec. 3). We continue with the data analysis, including calculating the network indices, forming groups of vulnerability levels, and checking for statistical correlations with development status (Sec. 4). The paper closes with a discussion of our findings (Sec. 5) and concluding remarks (Sec. 6).

13.2 Related Work and Research Gap

The degree of dependency on critical communication infrastructure is unevenly distributed across the globe. Already since the early 2000s, research on shortcomings and imbalances in the provision of internet access and the unequal exploitation of economic gains for societies was developed within the framework of the digital divide theory. In the early phase of broadband deployment from 1995 to 2005, the digital divide was explained by the presence and quality of physical access to the internet (W. Chen & Wellman, 2004), followed by increased research on the micro-level of digital skills and usage from 2002 onwards. In the third and nascent phase, attention is paid to outcome aspects and path dependencies of internet use (Van Dijk, 2017). In the following study, we want to combine features of the purely physical focus of the early research with the outcome side of increasing societal and economic dependency on the internet on global level. The growing demand for internet bandwidth throughout the COVID-19 pandemic has further raised awareness on these dependencies in 2020 and beyond (Feldmann et al., 2020; TeleGeography, 2021).

In the current academic research on the vulnerability of the submarine fiber optic network, several contributions approach the subject from the perspectives of diverse scientific disciplines (see Table 13.1). Various types of empirical academic case studies on specific countries (Muneez, M. et al., 2017; O'Malley, 2019), continents (Cariolle, 2018), and regions (Gerlach, C. & Seitz, R., 2013; Hummelholm, 2019; ITU Publications, 2018; Sutherland, 2009; Thorat, 2019) are presented, each with a distinct understanding of vulnerability. The studies of national focus incorporate a wide range of local proponents like geographic, geopolitical, and environmental context but do not offer definitions of vulnerability measurable through empirical means. In contrast to that, Cariolle defines digital vulnerability in his study on sub-Saharan Africa *"as the risk for a country and its population of its access to telecommunication services being hindered by failures in its telecommunication networks"* (Cariolle, 2018), taking into account internal digital divides of the 46 countries he analyzes. Whereas Cariolle considers the local perspective of countries as a single reference unit, Omer et al. evaluate the vulnerability of the global network *"by identifying the links in the network that would lead to greater damage than others when disrupted"* (Omer et al., 2009), putting more emphasis on the critical links than on the nodes. A similar edge-bound perspective is taken in the study of Palmer-Felgate and Booi on different SCC system designs. Here, vulnerability is understood in a broader sense as the absence of resilience, with the latter being statistically modeled through the availability of alternative routes, short repair times, and reliability of an SCC (Palmer-Felgate & Booi, 2016; Palmer-Felgate et al., 2013). Their study focuses on the edges of the submarine network, not the consequences of failure for the nodes. Furthermore, the routes used in their simulation omit several extensive sections of coastline, such as Australia, Oceania, Central America, and Sub-Saharan Africa (Palmer-Felgate & Booi, 2016). Within our work, however, we put

focus on the vulnerability in terms of internet access security from the perspective of redundancy availability in each territory. We only consider the meta-structures of the whole network to familiarize the reader with the global SCC network.

Study	*Perspective*	*Unit*	*Academic Discipline*
O'Malley, 2019	National	State	Security studies
Muneez, M. et al., 2017	National	State	Environmental studies
Hummelholm, 2019	Regional	Cities	Cybersecurity studies
Sutherland, 2009	Regional	States	Economics
ITU Publications, 2018	Regional	Islands states and overseas territories	Economics, development studies
Cariolle, 2018	Intra-continental	States	Development studies
Palmer-Felgate and Booi, 2016	(Partly) Global	Cable routes	Engineering
Omer et al., 2009	Global	Continents	Engineering
Research gap	Global	States and overseas territories	Critical infrastructure research

Table 13.1: Overview of empirical academic studies analyzing the consequences of submarine cable failures.

With the exception of the contribution by Omer et al. (Omer et al., 2009), the studies presented above have not reached the perspective of a global comparative analysis of the statistical population. As the exponentially increasing development of SCC numbers, bandwidths, global cable length, and internet traffic demand has continued since the years of data collection (2006, 2008) and publication (2009) of Omer's work, it is worthwhile to look at the present status of the submarine cable network more than a decade later. In addition, the capacities of the relevant statistical programs have further developed over the past decade, which allows a more detailed global view of the backbone structures than a rough comparison of the world regions.

Although Bischof et al. (Bischof et al., 2018) also take a global perspective, they focus on the consequences of higher latency in data traffic rather than on the total loss of internet connectivity in a country when an SCC is lost. The shortest possible latency can be crucial for certain sectors of the economy, such as high-frequency trading in modern finance. However, latency plays a minor role when viewed from the perspective of national internet supply vulnerability.

Consequently, there is a research gap in the global analysis of internet supply security, which considers all coastal and island territories as the central unit of analysis at the same time. To assess the necessity of preventive protection measures, it is important for legislators and authorities to compare the global internet infrastructure redundancies. Hence, the first goal of this work is to provide an up-to-date picture of the global internet supply security situation through SCCs for each autonomously regulated territory. The second goal is to examine whether the global digital divide is also reflected in SCC fault vulnerability.

13.3 Method

To approach both goals in a methodologically thorough way, we first find suitable definitions for the core concepts of this work (13.3.1). We then introduce the network analysis method and discuss the applicability of various centrality measures to the SCC network (13.3.2). Afterward, the specific scenario formation for the vulnerability estimation for the territories (13.3.3) is described. Subsequently, we present the availability (13.3.4) and compilation of the data sets (13.3.5) and discuss the network analysis software (13.3.6).

13.3.1 *Definitions*

The concept of vulnerability is a matter of considerable controversy in risk research. It is variously defined, depending on the context of use and the application of the psychological, social, or technical perspective (Feldmann et al., 2020). Therefore, this concept, which is central to the following work, needs to be discussed. Simply put by the Society of Risk Analysis, the core of most vulnerability definitions is "the degree to which a system is affected by a risk source or agent" (O'Malley, 2019). On the other hand, the UNDRR, in its own definition of vulnerability, specifies the context variables that influence the degree of vulnerability: "The conditions determined by physical, social, economic and environmental factors or processes which increase the susceptibility of an individual, a community, assets or systems to the impacts of hazards." (UN, 2016). The latter definition is adopted in the following, as it provides a comprehensive understanding of possible influencing factors. In this work, we will limit ourselves to the referent object of the community, which is located within the territorial boundaries of a state or an overseas territory. The influencing factors are primarily physical, as the connection to physical infrastructure is examined. It is worth noting that the UNDRR's definition of vulnerability considers individual, communal, and systemic levels of analysis as a reference object. This study applies the definition to states since they represent an applicable and appropriate level of analysis in a global comparison. In this perspective, a state fulfills two roles: First, it is an object of risk that is threatened by the impairment or loss of data traffic; secondly, by taking preventive measures and exercising its regulatory competence in terms of internet governance, it also influences the quality of a threat. Although overseas territories, in some cases, do not own full autonomy in terms of these competencies, they will be considered at an equal level, as it would not make sense in this context to consider them as part of their respective mainland. For example, it would contradict the purpose of the study to add the SCC connections and bandwidth of Martinique, Mayotte, and French Guiana to France, because all units are located in different geographical contexts. Consequently, for the purpose of this study, the geographical location is more important than the state affiliation of a particular region. Fortunately, the International Telecommunication Union (ITU) offers databases for each member state as well as overseas territories (see 13.3.5). The fact that internal inequalities regarding digital vulnerability within states and societies exist is hence ignored in favor of a globally quantifiable unit of analysis.

This study is based on the hypothesis that the number of cable connections in a territory and their transmission capacity (bandwidth) measures the extent to which a country

depends on an SCC's functioning. A low number of links to the global fiber optic network then indicates high vulnerability. Conversely, following the hypothesis, a comparatively high number of connections to the physical backbone infrastructure suggests many redundancies and thus low vulnerability. For this study, we define redundancy as an alternative, secondary infrastructure which provides the same or similar service as a potentially failed primary infrastructure. If multiple infrastructures provide this option, the plural form 'redundancies' is applied.

13.3.2 *Network Analysis*

The global internet backbone constitutes a complex network of relationships between a large number of coastal states and territories, whose connectivity is characterized by their position in it. Consequently, the fiber optic network can be analyzed not only as a physical network of fiber optic cables but also as an abstract network from which characteristics of its components can be deduced. With the method of network analysis, graph theory offers a tool that considers multidimensional group contexts and allows conclusions to be drawn from the position of nodes in the network and the connections (edges) between the nodes (Faramondi et al., 2020). Quantitative network analysis deals with the relationship of nodes by assigning quantifiable values. The presence of a communication cable (binary) as an edge and a metrically scaled property (e.g., their bandwidth) between two states or overseas territories are both quantifiable and can therefore be modeled using quantitative network analysis.

In the following, we will explain the conceptualization of the graph since certain features may limit the choice of centrality measures (Oldham et al., 2019). First, the edges will be undirected because the SCCs send data in two directions. Second, as there is the possibility of multiple SCCs connecting the same nodes, the network will be modeled as a multigraph. Replacing multiple edges with the cumulated weight in a single edge between two nodes is not possible, as we consider the bandwidth of the cables as the weight of the individual edges. Third, loops will not be formed because this would contradict the requirement to exclusively include *inter*national SCCs. The network is constructed through the graph G

$$G \quad V, E, r \tag{Eq.1}$$

with the nodes V, the edges E and the incidence function r defined as

$$r \quad E \rightarrow \{\{v, w\} \quad v, w \in V\ ,\ v \neq w\} \tag{Eq.2}$$

with v and w as distinct nodes potentially connected through multiple edges.

In general, network analysis offers measurements to assess the general network properties, demonstrating topological characteristics as well as different types of measures for centrality and efficiency of transmission. To familiarize the reader with the network, we have chosen $|V|$, $|E|$, largest component, maximum degree, edge density, mean distance, diameter, and largest clique as measures for an overview.

A wide array of global and local centrality measures can be applied to determine the positions of nodes – either in the global network or locally within their closest neighborhood (Oehlers & Fabian, 2021; Stergiopoulos et al., 2015). Since the network has a moderate size, the global network measures can be calculated for the graph. With a view to the second goal of this study, we want to form groups of highly vulnerable nodes. In this regard, those located in the peripheries of the global network, far from high values of centrality, are of particular interest. Most studies on networks fall under the paradigm of criticality, where highly connected central nodes are presumed to be of higher importance for the functioning of a network as a whole (Oldham et al., 2019). Considering the position of the weakest connected nodes is a somewhat atypical perspective.

The most frequently applied measures are degree centrality, betweenness centrality, and closeness centrality. For weighted graphs, strength and local efficiency are also often considered.

First, degree centrality *Degv* measures the number of connections to a node *v*, irrespective of their weight. It is a local measure, as the global network does not need to be known; it only counts the number of edges to adjacent nodes *w*. This is given through an adjacency matrix a_{vw}, whose elements take the value 0 if *v and w* are not connected by an edge and otherwise the value 1. The maximum degree of *G* is denoted as ΔG.

$$Degv \quad \sum_{w} a_{vw} \tag{Eq.3}$$

As the analysis is aimed at the availability of redundancies, degree centrality seems to be the obvious measure to apply. Firstly, it counts the edges, which is the class of components we identified as the independent variable. Secondly, as a local measure, it offers a simply calculated yet granular level of analysis. However, degree centrality does not consider the weights of the edges.

To account for weighted graphs like *G*, Barrat et al. extended degree centrality to vertex strength *Strv* (Barrat et al., 2007), as the sum of the weights of the local edges of a node. In our model, this corresponds to the local sum of the SCC bandwidths *b* adjacent to any given node *v*.

$$Strv \quad \sum_{w1}^{V} a_{vw} b_{vw} \tag{Eq.4}$$

A third measure, closeness centrality *Clsv* is determined by the average of the shortest paths δ of a node *v* with every other node of the network. High values of closeness centrality mean close relationships with many nodes.

$$Clsv \quad \frac{1}{\sum_{w \in G} \delta v, w} \tag{Eq.5}$$

Based on the closeness centrality, Hao et al. found that balancing node traffic in network design can enhance robustness regarding cascading failures (Hao et al., 2020, 2021).

Fourth, another measure based on shortest paths is betweenness centrality *Btwv*. It reveals the frequency of node *v* being a transmitter of information in the network. This is achieved by dividing the number of all shortest paths of any other nodes in *G* in which node *v* is present with all geodesic distances δ in *G*. For example, betweenness has been applied by Nguyen et al. on attack strategies on networks to identify those nodes, which removal would lead to longer δ, making information exchange more costly (Nguyen et al., 2019).

$$Btwv \quad \sum_{u \neq v \neq w} \frac{\delta_{u,w}v}{\delta_{u,w}} \tag{Eq.6}$$

Finally, the local efficiency $Eff_{loc}v$ measure first introduced by Latora & Marchiori (Latora & Marchiori, 2001) for small-world networks was later modified to extend its application to complex networks with weighted and multiple edges (Latora & Marchiori, 2003). It quantifies the fault tolerance of the immediate neighborhood of *v* to cope with the removal of *v*. Therefore, G_v denotes the subgraph of the neighborhood of *v* without *v* itself and $EffG \quad \frac{1}{VV-1} \sum_{v \neq w \in G} \frac{1}{\delta v,w}$.

$$Eff_{loc}v \quad \frac{1}{V} \sum_{v \in G} EffG_v \tag{Eq.7}$$

We are not modeling global internet outages but the consequences of single, double, or triple cable failure on a territorial level. Hence, we limited the choice of centrality measures to those offering benefits in the perspective of edge removal in *G* (see 13.3.2). For the SCC network, it must be considered that single nodes may be connected to only one other node but by several parallel edges. If measures have the shortest path as the basis of their calculation, removing an edge does not change the values if a parallel edge connecting the same nodes replaces it. Consequently, closeness and betweenness centrality are not producing meaningful results with multigraphs. Therefore, we rejected closeness and betweenness centrality as suitable measures for redundancy analysis. Local efficiency works with the simulation of a node failure, which contradicts the idea of measuring SCCs as edge failures. Nagurney and Qiang modified local efficiency into a network performance measure to be applied to both components of a graph (edges and nodes), but only for directed graphs (Nagurney & Qiang, 2007). As *G* is undirected, the modified local efficiency is not further considered.

To support the decision for an adequate redundancy metric, a basic assumption for state backbone access needs to be kept in mind. The total available bandwidth of node *v* (B_v) is the sum of edge bandwidths *b* from the available points of access to external networks, be it through SCCs or LCCs:

$$B_v \quad b_{SCC\ v}\ b_{LCC\ v} \tag{Eq.8}$$

Applied to an SCC network, degree centrality and strength as redundancy measures both have a key disadvantage. Degree alone does not consider the highly diverse bandwidths of SCCs, ranging from 1 GB/s to 250 TB/s. On the other hand, the strength measure can only be applied for SCCs, as there is no available data for cross-border LCC bandwidths. Satellite communication is, due to its low bandwidth and low prevalence in the population – rooted in high prices and long latency time – not considered an equivalent redundancy and is therefore omitted from the analysis. To overcome the problem of only partially available bandwidth data, we have decided to perform a two-step calculation. The first step is to cluster groups based on the number of backbone accesses and is intended to identify potentially vulnerable units. SCCs and LCCs are treated equally in the failure scenarios applied for the formation of groups (see 13.3.2). In the second step of analysis, the share of individual local SCC in the total SCC bandwidth (strength) of a node is measured. This is necessary to account for the large spectrum of bandwidths of globally installed SCC. The local proportion p of the weight of an edge y to the $Strv$ can be modeled accordingly as percentage:

$$p_{b_y}v \quad \frac{b_y}{Strv} \times 100 \qquad \text{(Eq.9)}$$

The higher $p_{b_y}v$, the more vulnerable a node is to the loss of edge b_y. A value of p_{b_y} 100 would mean that cable y is the only access to international networks. Vice versa, the closer the value approaches 0, the less the contribution of a cable to the connectivity of a territory is to be rated, which is why the state's vulnerability is also reduced for the potential loss of this individual cable. The local weight of all adjacent SCCs is being calculated for each territory. Subsequently, the cables can be arranged by their locally weighted capacity input, enabling the application of worst-case scenarios.

13.3.3 *Scenarios and Group Formation*

Information and communication technology (ICT) infrastructures are subject to different influencing factors during times of peace, crisis, or conflict (Reuter, 2019). We applied the classification of Aceto et al., 2018 on various SCC disruption events which we then formed into corresponding scenarios (see 13.1). To simulate local SCC disruption scenarios, the consequences of removing the edges with the highest bandwidth in the overall strength of the node under consideration are examined. Therefore, we apply three scenarios, in which each state is simulated to lose its first ($S1$), the top two ($S2$), and the top three ($S3$) data-carrying edges.

With an average of 100 yearly incidents, single SCC disruptions are common (Mauldin, 2017), making the occurrence of $S1$ by far the most probable. Unintended human incidents like anchoring and fishing accidents triggering cable rupture typically led to the single SCC loss scenario $S1$. A variety of situations exist that can potentially result in the simultaneous loss of several cable connections, e.g., a seaquake and subsequent underwater landslides. This kind of cascading incident has led to multiple cable breaks in various regions in the past (Palmer-Felgate et al., 2013). There have been cases of multiple cable losses parallel in time, as in the Egyptian incident of 2008 and Taiwan's situation in December 2006 (Carter et al., 2009; Kitamura et al., 2007). With $S2$, we

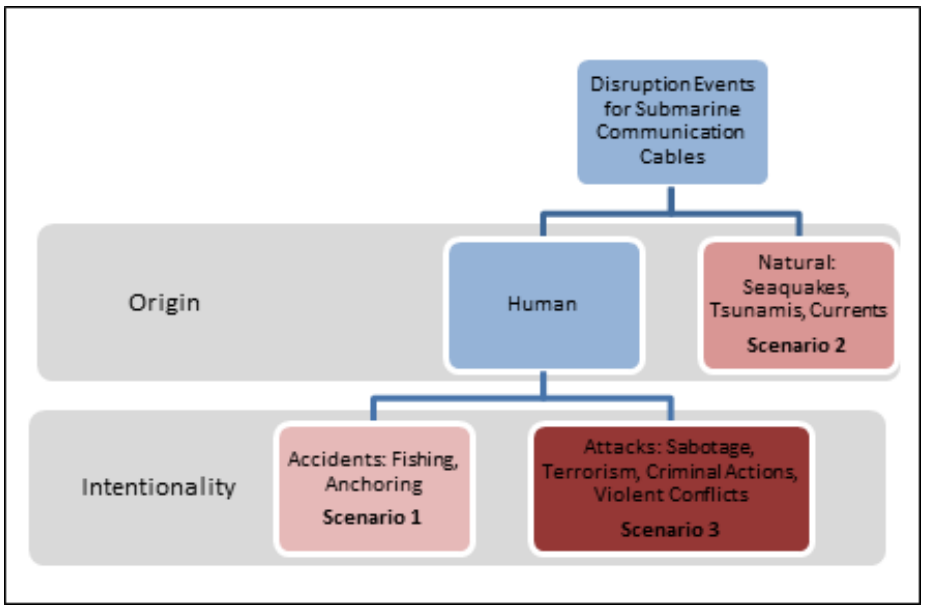

Figure 13.1: Classification of disruption events and corresponding scenarios for submarine communication cables (Own representation based on the category system of (Aceto et al., 2018)).

intend to model these incidents of parallel small-scale SCC disruptions. Meanwhile, there are no criminal, terrorist, or military interference records with SCCs or cable landing stations (CLS). However, targeted attacks like sabotage bear the potential for multiple simultaneous cable connection losses if conducted with accurate timing (Calle et al., 2019; Rueda et al., 2017). As sabotage, terrorism, and criminally motivated actions will likely target the edges of the highest bandwidth, we hold $S3$ to be an appropriate reflection of a multiple-loss scenario with coordinated targeting of the top edges. By reducing our focus exclusively to physical disruptions, we do not consider those outages triggered by governmental interference with internet traffic or the different types of cyberweapons (Freyburg & Garbe, 2018; Reinhold & Reuter, 2021). Since the cables themselves do not contain any software components, it is more appropriate to model these types of disruptions with the removal of other, land-bound nodes.

The local scenario $S1$ for node v is being modeled as

$$S1_v \quad \frac{b_{\max y}}{Strv} \times 100 \qquad \text{(Eq.10)}$$

where $b_{\max y}$ is the SCC with the locally highest bandwidth in v. This allows us to calculate the proportion of criticality for the strongest SCC. The scenarios $S2$ and $S3$ are modeled accordingly:

$$S2_v \quad \frac{b_{maxy}\ b_{\max n-1y}}{Strv} \times 100\,; \qquad \text{(Eq.11)}$$

$$S3_v \quad \frac{B_{maxy} \; B_{\max n-1y} \; B_{\max n-2y}}{S\,trv} \times 100. \qquad \text{(Eq.12)}$$

The units are divided into three groups according to the scenarios: Units that have no redundancies, meaning their broadband connection depends entirely on the operation of one SCC, are assigned to Group 1. Units that encounter a complete failure of their broadband connection within the given scenarios $S2$ or $S3$ are assigned to Group 2. All other nodes – having connections to more than four SCCs or LCCs in sum – are assigned to Group 3. Within the groups, rankings are identified according to specific group characteristics (see 13.4.2).

13.3.4 *Availability of Data Sets*

There are three comprehensive compilations of the worldwide SCC paths regarding the data sets. While both data sets provide information on cable names, approximate cable runs, adjacent CLS, length, and operational status, they differ in terms of additional information and the number of cables listed. First, the *Submarine Cable Almanac* (*SCA*) of the Submarine Telecoms Forum (SubTelForum, 2020a) provides a list of global cable routes (n=301), supplemented with information on the transmission capacity of the cables. The report is updated quarterly with publicly available data from the submarine cable industry. Second, more detailed map material (n=480) is provided by the *Submarine Cable Map* of the online platform TeleGeography, where the owner and operator companies are also listed (TeleGeography, 2020). Still, the bandwidths of the cables are not specified (TeleGeography, 2020). Third, the *Infrastructure Map* of the online capacity marketplace Infrapedia also offers detailed SCC capacity data submitted by experts and constantly validated (Infrapedia, 2023). In the rare cases of conflicting information on the properties of an SCC, we incorporated the information from the *SCA* to prevent potential entry errors of the *Infrastructure Map*, which has only been in operation since 2019. Regional data sets, such as the map of the African Network Startup Resource Center (AfTerFibre, 2023), were used to verify the information. The validation of the bandwidth data can be done for cables that are in operation until 2016 with data from *Greg's Cable Map* (Mahlknecht, 2016). ITU's *Interactive Transmission Map* (ITU Publications, 2020) was used and validated with data from the *Infrastructure Map* for data on the quantity of cross-border LCCs. The information of the maps mentioned above is publicly available for non-commercial use, and the *SCA* is accessible online and free of charge (AfTerFibre, 2023; Infrapedia, 2023; Mahlknecht, 2016; SubTelForum, 2020a; TeleGeography, 2020).

The ITU surveys the international broadband traffic for each member state, with the information given directly by governments through yearly questionnaires. The data is provided in the *World Telecommunication/ICT Indicators Database*, which is updated every six months. It contains 180 measures of 200 countries, including "*International internet bandwidth, in Mbit/s*", "*Lit/equipped international bandwidth capacity*", and "*International bandwidth usage*" (ITU Publications, 2019b). The database also contains information on various internet usage indicators in the population (ITU Publications, 2019b). The *World Telecommunication/ICT Indicators Database* is only available on payment basis.

13.3.5 *Data Compilation*

Since the intention is to assign properties to edges and nodes in a network, two data sets required to create the model are introduced in this section. On the one hand, the edge list is crucial, because the connections between the units and their properties are listed there. On the other hand, the node list describes specific properties of the connected units, the states and overseas territories.

Compilation of the Edge List

Three categories of submarine cables were excluded from the data set and thus omitted in the model. Firstly, cables whose CLS are located within the same territory were not included because they do not contribute to the international data traffic of a state/overseas territory. Examples are *ADONES*, the domestic Angolan submarine cable network and the *JaKa2DeLeMa* intra-Indonesia cable system (SubTelForum, 2020a). The second exclusion category consists of cables only intended for data use at sea, such as oil drilling platforms or shipping. An example of an offshore system is the *TampNet* system installed in the North Sea (SubTelForum, 2020a). Lastly, SCCs that connect military bases and do not contribute to the local connectivity are also excluded from the sample. The only two occurrences in the SCA are the *GTMO-1* and *GTMP-PR*, connecting Guantanamo Bay US Naval Base with Florida and Puerto Rico.

This leaves 197 active cable systems out of 301 cables listed in the SCA, which are modeled over 605 edges between territories. We omitted those nodes adjacent only to excluded SCC. While the majority of these SCCs have two CLS, more complex cable systems also exist. With 33 connected units, *SEA-ME-WE 3* has the highest global connectivity for a single cable system. Cable systems like these are represented in the model by individual edges between territories, as in the case of *SEA-ME-WE 3* by 32 individual edges.

Numerous variables are integrated into the edge list. Essential variables are the states and overseas territories where an SCC is landing. In addition, there is the bandwidth as a measure of the weight of the cable, which is coded in terabytes per second (TB/s). In public data sets, the design capacity is usually given as the maximum capacity of a communication cable expected at the time of design. However, older cables can be used far beyond their original design capacity by applying new technologies like wavelength-division multiplexing (Hadaway et al., 2016). If there is an upgrade for the capacity of a cable that exceeds the design capacity, the upgrade capacity is applied.

Other variables include the location of the CLS, the years of commissioning and expected end of operation, length in km, ownership, and construction costs. These are not necessarily used in the model but have been integrated into the data set for advanced data visualization or subsequent research projects.

Compilation of the Node List

In the node list, the specific data on coastal states and overseas territories are combined. The node list only includes states and territories considered in the edge list. This leaves coastal and island units without qualified SCC connections out of the analysis, for example, East Timor, Poland, or Slovenia. This limitation leads to a reduced sample, including 169 states and overseas territories for the following analysis. This figure may change in the future as more territories are connected to the submarine cable network. For example, with the completion of *Southern Cross NEXT,* Kiribati and Tokelau will be connected by submarine cable for the first time (Qiu, 2020).

For the node list, essential variables for the analysis can also be distinguished from auxiliary variables for better visualization. Essential variables are the name of the respective state or territory, the number of cable accesses (credit), the sum of the bandwidth of the SCC connections to a unit (strength), and the number of alternative internet resources. The latter is composed of the number of adjacent cross-border fiber optic LCCs that were counted on the basis of the ITU Interactive Transmission Map (ITU Publications, 2020).

Other variables that serve to visualize the data are the geographic data of the territories. We used the geographical center (centroid) of units for simplicity, which we took from the *rworldmap* expansion program in *R* (South, 2011). To be able to test the hypothesis of the digital divide, the socio-economic development status of the states and overseas territories from the M49 Standard of the UN Statistics Division is included for every unit (Division, 2018).

13.3.6 *Statistical Analysis Program*

The network analysis is performed with *R (Version 4.0.0)*. With the packages *igraph* and *sna*, two libraries are available to execute a social network analysis (Butts, 2019; Csardi & Nepusz, 2006; McCulloh & Perrone, 2017). Since *igraph* offers more functions than *sna* and performs faster for networks with over 150 nodes (Butts, 2019), we modeled the network with *igraph (Version 1.2.6)*. Another argument in favor of the *igraph* package is that interactive visualization can be carried out using an *RShiny* web application via the extension package *igraphinshiny* (Lee, 2016).

13.4 Findings

A purely visual analysis of the global network offers limited advantages due to the quantity of nodes and edges, as the network model overlaid on a world map in Figure 13.2 illustrates. Therefore, a resort to mathematical network parameters to identify the structures in the network is necessary (Table 13.2). The modeled graph consists of 169 units (states/territories) connected by 613 edges (submarine cables and cable system branches). Each node is part of the largest component so that there is at least one possible path between each node in G through which information can be exchanged, making G a connected graph.

The edge density of the model is 4.31%. This indicates the ratio of the actual edges to the number of possible edges. Thus, the submarine fiber optic network is relatively loose. The mean distance between two nodes is 4.44 edges, while the network's longest possible distance (*diameter*) consists of nine edges. A *clique* is a group of several nodes in which each member has at least one direct edge to every other member. In G, the largest clique of five members exists between the Southeast Asian states of Thailand, Indonesia, Hong Kong, Malaysia, and Singapore. A dense network of submarine cables connects these territories. In addition, there are another 18 cliques of four units each, in East Asia, the MENA region, Southern Europe, and the Caribbean.

Measure	*Value*
$\|\mathbf{V}\|$	169
$\|\mathbf{E}\|$	613
Largest component	169 (all nodes)
$\mathbf{\Delta G}$	58
Edge density	0.0431
mean distance	4.44
diameter	9 edges
largest clique	5 nodes

Table 13.2: General properties of the network model.

After this short overview on the properties of the network as a whole, this section continues with node-specific evaluations through the identification of central nodes (13.4.1), the formation of groups of redundancy levels (13.4.2), and the testing of the hypothesis of the Global Digital Divide (13.4.3). Subsection 13.4.4 summarizes the findings of the previous sections and merges them into the overall result.

13.4.1 *Identification of Central Nodes through Centrality Indices*

When examining individual nodes, the centrality measures offer a way of identifying particularly central and marginal nodes (Marzo et al., 2018). The most straightforward measure of centrality is *degree* centrality. For this purpose, each territory is examined locally for the number of its adjacent edges. With 58 adjacent SCCs, the USA clearly leads the ranking, while the United Kingdom and Japan dominate their regions with a degree centrality of over 30. Egypt's high value can be explained by its role as a transit country for SCCs. Since the shortest sea connection between Europe and East Asia leads through the Suez Canal, a large number of SCCs – parallel to shipping routes – run through the Egyptian mainland. The high values of Italy and Saudi Arabia also reflect this effect, as both countries are often the next CLS after the Egyptian bottleneck. With Hong Kong, Singapore, and Malaysia, there are also three East Asian trade and technology centers in the top ten of degree centrality. At the lower end of the scale, 15 units are connected to only one SCC. At first glance, these can be divided into two categories: On the one hand, Northern and Eastern European countries may be provided with sufficient bandwidth through fiber optic land connections to allied states (e.g., Croatia, Lithuania, and Romania). On the other hand, there are island states and territories that are connected to the submarine fiber optic network without redundancy due to their geographical isolation, e.g., the Marshall Islands, Palau, and the Seychelles. Furthermore, 28 territories can be identified with only one redundancy, i.e., two SCC in sum.

13.4.2 *Assessing State Vulnerability to SCC Loss through Fault Scenarios*

To account for different levels of redundancy, the units are classified according to the group formation as described in section 13.3.2. For each of the three groups, a short introduction is followed by an exemplary case study to facilitate the interpretation of data. We chose a distinct visualization of single critical SCC sections (group 1) or the simulations of failure scenarios (group 2 and 3) for each group.

Group 1: No Redundancy

The first group includes units without redundancy in case of failure of their single SCC – neither SCC nor LCC. Figure 13.3 lists these 15 units along with the SCC through which they are being supplied. Included are territories that are connected by a short cable system branch, such as Gibraltar. The distance from the T-junction of the *Europe India Gateway* system to the CLS in Gibraltar amounts to only 15 km, where the worst case of a complete loss of bandwidth can occur. On the other hand, New Caledonia is dependent on a single cable – rather than a multi-station system – to connect with Australia. The distance over which damage to the cable named *Godwana-1* would be critical to New Caledonia amounts to 2150 km.

The exemplary case study concerns the Seychelles, a state whose access to the global network depends on a single SCC, the *Seychelles to East Africa System*, linking the

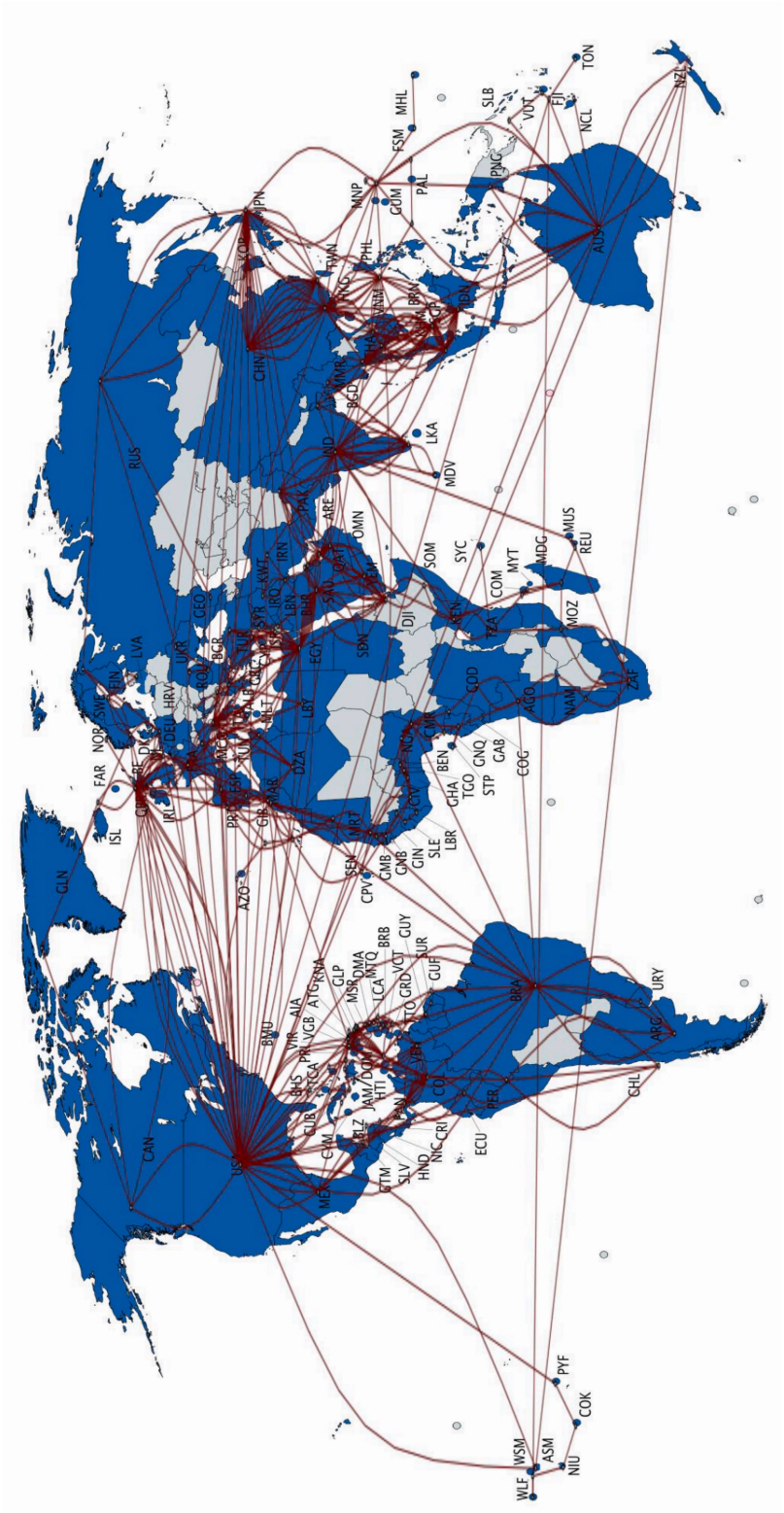

Figure 13.2: Visualization of the global SCC network (red) between all qualified units (blue), named with their ISO3 country code (Own representation through igraph (Csardi & Nepusz, 2006) and mapchart).

capital Victoria on the main island Mahé with Dar es Salaam in Tanzania. Island states and island territories in group 1 cannot obtain LCC access, which implies that S1 already leads to a 100% loss of potential broadband connectivity. The critical cable distance of an SCC failure that would amount to full connectivity loss is approximately 1811 km. The *Seychelles to East Africa System* was commissioned in 2012 with a total capacity of 320 GB/s. This potential bandwidth exceeds the actual demand for bandwidth (2018: 4.2 GB/s) by a multiple, which is partly due to the medium level of internet usage among the population (2017: 58.11%) (ITU Publications, 2019b) and the relative novelty of the cable, which means that its end-of-service date is not planned until 2037 (Infrapedia, 2023). To remain commercially viable for a quarter of a century, the design bandwidth capacity of modern cables usually far exceeds the demand of a unit in the ready-for-service year. In the case of Seychelles, a projected second backbone connection is planned through a branch of the *Pakistan & East Africa Connecting Europe (PEACE)* cable system in 2021. This development can be assessed as positive from the perspective of the Seychelles. Yet, the PEACE cables system as a whole is also the subject of geopolitical debate due to the involvement of Chinese companies in its construction (Fouguet, 2021).

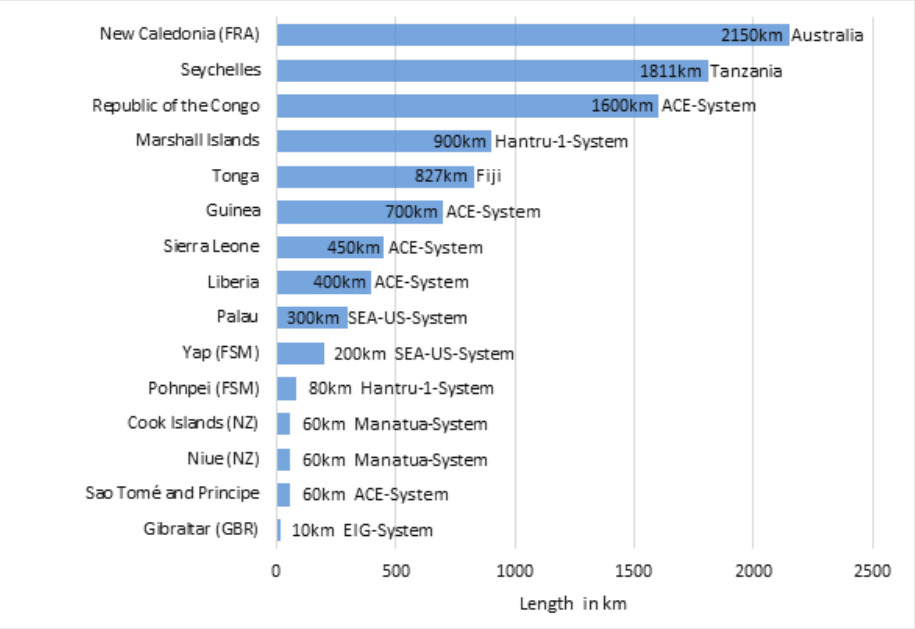

Figure 13.3: Length of critical cable sections of units with single SCC connection and missing LCC redundancy (Group 1) with the adjacent SCC system or connected territory (own figure).

Group 2: Low Redundancies with Balanced and Imbalanced Fallback Levels

The units that suffer 100% connection failure in S2 or S3, i.e., units with a maximum of three SCC connections, are classified as the second group of units in this study (Figure 13.4). To additionally take the potential of redundancy through LCC connections of coastal states into account, only units with less than three cross-border LCCs are

considered to belong to this group. It is advantageous for this group of 19 units to have a balanced fallback bandwidth ratio to maintain connectivity and minimize risks if the widest and second widest cables fail. Iceland can be considered a good example in this respect. Despite its insular situation, the unit is served by three different cables, all transmitting a fairly similar amount. Even with the removal of the widest cable (*Greenland Connect*), 58.73% of the bandwidth is maintained by the capacity of the other two cables (*Farice-1* and *Danice*).

In contrast, the ratio of Uruguay's cables is rather unbalanced. There, the elimination of the *Tannat* SCC would result in a relative bandwidth loss of 95.74%, with only 2 TB/s weighted design capacity remaining for the whole of Uruguay through the *Antel* cable. As a case study, Samoa is chosen, which has connections to two cable systems: The *Manatua One* and *TUI Samoa.* Both SCCs deliver a combined weighted capacity of 8.5 TB, with 6 TB/s accounted for by Manatua One and 2.5 TB/s by TUI Samoa. This results in a 70.59% to 29.41% ratio of potential bandwidth capacity. Although the fallbacks are not ideally balanced (50%/50%), a sufficiently balanced redundancy level can be assumed, especially considering the equipped international bandwidth capacity of only 4 GB/s (ITU 2017). This leads to a problem arising from the similar CLS in Samoa's capital Apia, which means that the main island has a single point of failure despite its SCC redundancies.

Figure 13.4: Scenarios of SCC failures for units with low redundancy in relation to their overall SCC connectivity, in the order of relative connectivity loss severity in S1 (own figure).

Group 3: High Redundancy

The remaining 126 units are assigned to group 3. This includes all units equipped with three or more redundancies, regardless of whether they are SCCs or LCCs. These units may suffer from a reduction in the data flow, for instance, due to reduced capacities. However, a complete territory-wide failure of the critical telecommunications infrastructure is unlikely. Further differences can be identified within this group, such as the sum of SCC and LCC connections. As point of reference, Figure 13.5 depicts the scenarios for the G20 member states. The USA is leading in terms of the availability of redundancy after the application of S3. Saudi Arabia holds the lead in S1, Japan in S2, while the USA remains among the top 3 of the G20 in both latter scenarios. Regarding the USA, excellent availability of redundancy can therefore be determined. The USA is leading the ranking of incident edges in the model with ΔG 58. Thus, the USA is considered a representative case for the group of states with high SCC redundancies.

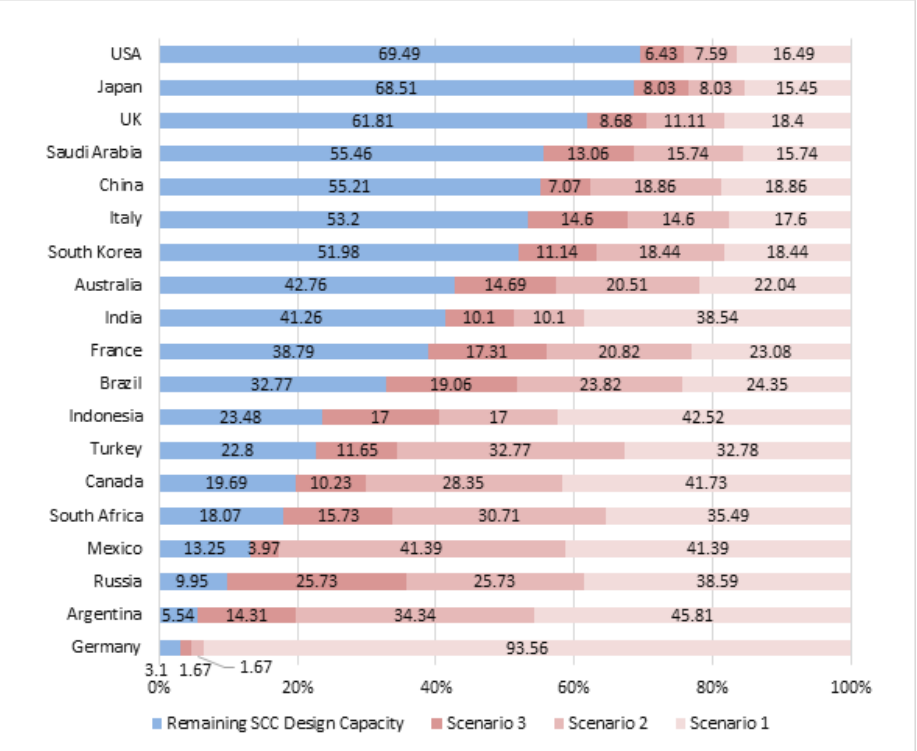

Figure 13.5: Scenarios of SCC failures for the G20 members in relation to their overall SCC connectivity. The European Union was omitted due to its status as an association of states. The number of LCC connections providing additional redundancies is specified in the right column. (own figure)

The potential international SCC bandwidth for the USA is estimated at 606.4 TB/s. The application of scenario 1, the connection loss of *MAREA*, results in a reduction of the potential SCC bandwidth by 100 TB/s. In relative terms, the failure of the widest cables would therefore amount to a 16.5% reduction of the overall potential SCC bandwidth. In scenario 2, in which the cable *BRUSA* additionally fails, the USA is missing 146.5

TB/s, respectively 24.1% of the potential SCC bandwidth. The loss of the three strongest cables in scenario 3 translates into a potential loss of 185 TB/s. Thus, even in the worst-case scenario simulated in this study, the USA would still have a potential remaining bandwidth of 421 TB/s. A comparison with the actual bandwidth needs of the USA (2017: 36 TB/s, estimated 2020: 42 TB/s) reveals conclusively that the USA's access to the global network will be maintained even under severe failures. In addition to its maritime connections, the USA is also equipped with LCC-Connections to Canada (n=22) and Mexico (n=9) and a vast access to the internet provided by satellites (CIA, 2021; Union of Concerned Scientists, 2022). Figure 13.6 visually sums up the group classifications on a world map while including the different types of omitted units as well.

13.4.3 *Socio-Economic Development and Redundancy Levels*

To test for the digital divide phenomenon, we combined our classifications of the analyzed territories with their respective socio-economic development status in Table 13.3 below. In the case of statistical independence between the socio-economic development status and the level of redundancy, the actual shares of redundancy groups within the UN M49 Standard subsamples would be close to the expected proportions derived from all territories in the second column of Table 13.3. However, we found large variation in group proportions within the socio-economic subsamples:

Of 39 units that we identified as developed countries, we assigned 34 (84.61%) to the highest redundancy group 3, four units (10.26%) to the medium redundancy group 2 and only one (2.56%) unit (Gibraltar) to the low redundancy group 1. Due to its geographical location and size, Gibraltar is an exceptional case. It has redundancy from the Spanish mobile network, at least in certain areas close to the border. Foreign mobile networks were not categorized as sufficient redundancy in the analysis, which consequently leaves the individual case of Gibraltar in group 1. Based on the data, we concluded that the geographic distribution of SCC connections in the Global North is reasonably good. This results in a very low probability of internet failure due to SCC outages in the aggregate, even in locations with multiple simultaneous outages (S3), by rerouting traffic through the available alternative routes.

Meanwhile, parts of the Global South are more at risk of experiencing a loss of broadband connectivity due to SCC losses. Accordingly, among the 106 developing countries that were not categorized as least developed countries (LDCs), 77 units (72.64%) were placed in group 3, 19 units (17.92%) were assigned to group 2, and ten units (9.43%) were listed in group 1. Among the total 24 LDC units, we allocated four units to group 1 (16.17%), which is a comparatively high proportion. In addition, five units (20.83%) are assigned to group 2 and 15 (62.50%) to group 3. Based on this matrix, the χ^2 value is 6.1015 with 4 degrees of freedom. As this study analyzes the statistical population of all coastal and island states and territories fitting into the pre-defined limits and no random sample has been taken in the process of analysis, no p-value is determined. With a value of Cramér's V = 0.27, it can thus be assumed that there is a moderately strong correlation between redundancy level and socio-economic development. Thus, the hypothesis of the Global Digital Divide can be confirmed for the availability of redundancies in developing countries and especially for LDCs.

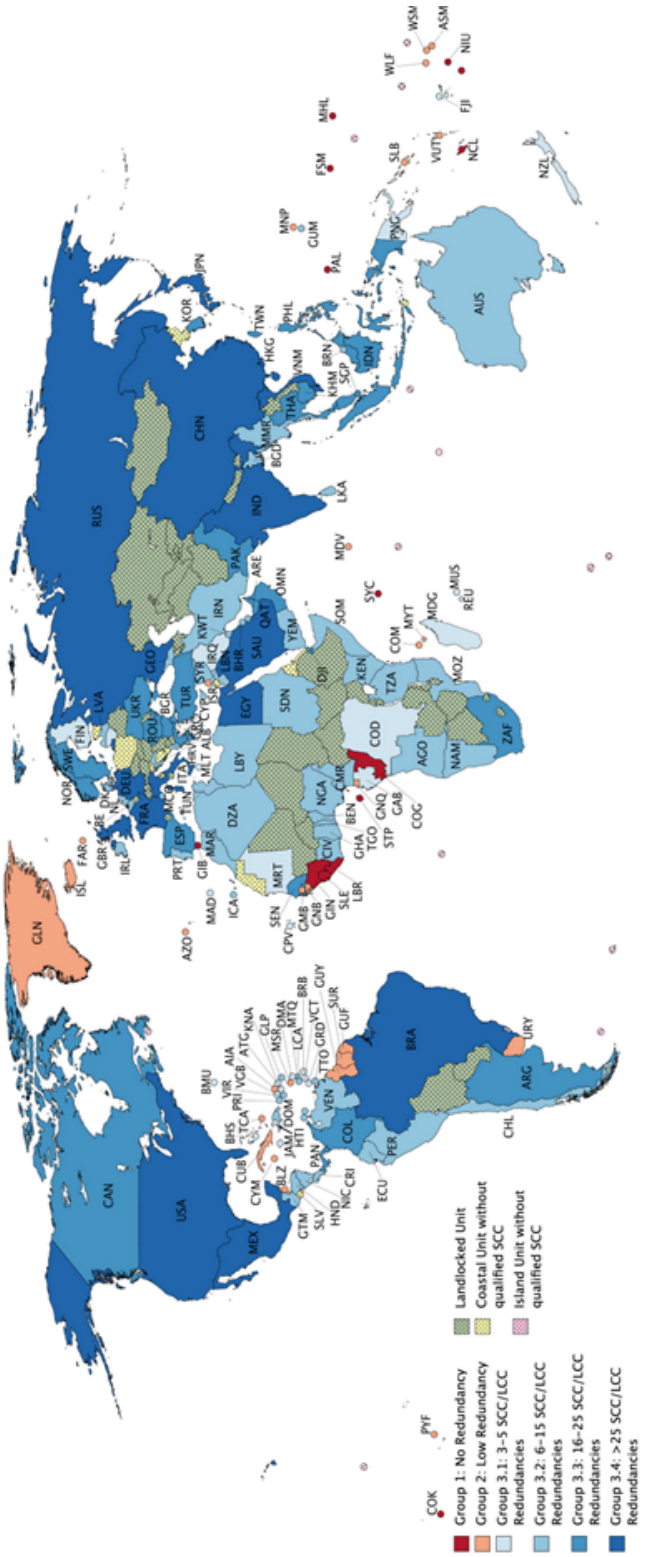

Figure 13.6: World map depicting classification of units into groups of high (blue) to low (red) redundancy through SCCs and LCCs. Landlocked units (green), units of coastal and island location without qualified SCC connection (yellow, pink) were omitted from the model. (own figure, created with mapchart)

	Socio-Economic Development Status (UN M49 Standard)							
Redundancy Level	***All Territories***		***Developed Countries***		***Developing Countries***		***Least Developed Countries***	
No Redundancy	15	8.88%	1	2.56%	10	9.43%	4	16.17%
Low Redundancy	28	16.57%	4	10.26%	19	17.92%	5	20.83%
High Redundancy	126	74.56%	34	87.18%	77	72.64%	15	62.50%
Sum	**169**	**100%**	**39**	**100%**	**106**	**100%**	**24**	**100%**

Table 13.3: Group classification in absolute and relative values and assigned to their socio-economic development status.

13.4.4 *Summary of the Findings*

In subsection 13.4.1, we revealed an imbalance with regard to the overall network structure. There is a wide variation in the different centrality measures. Yet, it is mainly the same – often developed – units topping the rankings in the positive sense in the indicators of node importance. Vice versa, some countries appear at the lower end of various centrality rankings in different centrality measures. This result is partly due to certain overlaps in the calculation of the centrality measures, yet the tendency of an unbalanced network can nevertheless be concluded from it. While this is not an original new finding, it represents a central role as confirmation of the propositions of older studies (Cariolle, 2018; Thorat, 2019) that the imbalance in the internet backbone still applies in the global context for the chosen time of analysis in mid-2020.

Keeping this fundamental observation in mind, subsection 13.4.2 aimed to precisely identify those units that, due to their position in the internet backbone, suffer total bandwidth losses (group 1, n = 15) or a significantly increased risk thereof (group 2, n = 28) due to an SCC failure. For group 2, it can also be noted that limitations in data traffic speed may also occur, especially in the territories with an unbalanced level of redundancy. On a positive note, the large majority of the units examined (n = 126) were assigned to group 3, thus assuming a sufficient level of redundancy for them.

Based on the group classifications we developed, we then determined the correlation of socio-economic development status with the level of redundancy in subsection 13.4.3. Among LDCs, there is a clear overbalance in terms of the group 1 status redundancy. At the same time, developed countries are more likely to be members of group 3 than the expected frequencies would predict. The hypothesis of the Global Digital Divide can thus be validated for LDCs and developed countries. Regarding the non-LDC Developing Countries, which make up the center group of the development status variable in the present study, we assigned some units to the lower redundancy levels. However, their occurrences are close to the expected frequencies or proportions of those within their development status group; see the fourth column of Table 13.3. These differences within the respective groups could best be explained by the inclusion of further variables but cannot be explained by our model.

13.5 Discussion

Both central variables chosen for analysis, grouping into three redundancy levels and United Nations Statistics Division (UNSD) socio-economic development status, can be discussed. First, classification based on fault scenarios provides the opportunity to make clear divisions into internally homogeneous and comparable groups. A metric measure would have hindered a concise intra-group representation of the relevant vulnerable units. Second, while widely used for academic studies with a global focus, the UNSD socio-economic development status has not been immune to criticism. The rather crude subdivision into developed, developing, and least developed regions creates very broad categories that are highly heterogeneous within. We chose this indicator anyway because it makes an objective classification of overseas territories possible. This puts it ahead of other indicators of socio-economic development, such as the Human Development Index. The fact that the UN uses the ternary variable in reporting of the Sustainable Development Goals underscores its lasting relevance. Due to the global focus, we decided to work with two ordinal variables in this study. Therefore, an analysis of the SCC backbone structure with applications of variables on ratio scale is advised for analyses of smaller scope, such as regional approaches or case studies.

The findings of this study must be seen in light of some limitations: First, states and territories are treated as a black box in this study, which masks local differences and discrepancies in failure resilience, for example, between rural and urban contexts, prosperous and poor communities, or household and corporate customers. This is rooted in the global focus of the study. An overly detailed level of analysis would have been at the expense of the clarity of the results. Second, we did not include landlocked states in the study. This can be explained by the lack of data on the design capacity bandwidth of LCCs. Moreover, by definition, landlocked states can only access the submarine fiber optic internet backbone passively via the transit of an adjacent coastal state. Consequently, landlocked states depend less on the functioning of an SCC than on the function of their LCCs and the willingness of their neighbors to forward internet traffic. However, a specific study on the status of landlocked states would be conceivable, and the study of Liu et al. (Liu et al., 2020) already offers a good starting point for further exploration.

Third, the actual usage of a cable (lit capacity) is rarely made public, creating a data gap. This study takes the perspective of supply security, therefore rendering this point irrelevant. Nevertheless, the analysis of the maximum (design) capacities is not suitable for telecommunication market analyses, as they do not provide any information on the share of cable utilization through leasing by telecommunications providers. This point is not so much a limitation of the analysis itself, but rather a reminder to interpret the data with appropriate caution. Fourth, the model disregards the varying risk of cable failure by applying the same scenarios to each country. Palmer and Booi discovered significant differences in the likelihood of cable failure depending on its geographic location (Palmer-Felgate & Booi, 2016). Correspondingly, high traffic volumes from fishing and cargo shipping, shallow waters, and tectonic activity in a maritime area raise the risk of SCC failure accordingly. Figure 13.3 should therefore be interpreted with caution since it has not been determined how dangerous the geographic contexts of the respective cable sections are. Future studies may combine our results with failure

probabilities to provide an even more realistic estimation of a state's vulnerability towards SCC failure.

Recalling the vulnerability definition of the UNDRR, it has to be kept in mind that the vulnerability of a reference object to a threat is not determined solely by the presence or absence of one aspect like redundancy. Instead, states can influence their individual vulnerability in the internet backbone cable system by preventive action like declaring cable protection zones. While these solely protect against accidental cable damage by establishing fishing boundaries and anchoring prohibitions, they do not protect against natural hazards. Avoiding geologically active zones on the seabed along the cable route is required to prevent destruction by natural hazards, particularly those triggered by tectonic activity. Regulative bodies may advocate this in negotiations with SCC installing and operating companies. Additionally, developing comprehensive reaction plans based on failure scenarios, the close-by stationing of repair resources (material, technicians, and vessels), as well as satellite internet receiving devices, would further lead to enhanced resilience against SCC failures. This study's results can thus help assess the redundancy level and, if necessary, justify the need for preventive measures.

Another aspect that needs to be investigated to better limit vulnerability is the consequence side of internet blackouts after SCC incidents and what measures have proven helpful in the event of damage. An index that globally compares the criticality of internet connectivity on the social, economic, or administrative levels does not yet exist. However, quantification of interactions between different critical infrastructures utilizing dependency risk graphs allows a better estimate of the consequences of a failure. By modeling potential cascading effects of critical infrastructures along dependency risk chains, cost-efficient mitigation strategies can be pursued (Stergiopoulos et al., 2015). With ongoing digitalization efforts in critical infrastructure sectors such as food, health and water, a growing complexity of these risk chains is likely (Katina et al., 2014; Kuntke et al., 2022; Thompson et al., 2019).

Moreover, van Eeten et al. have found that energy and telecommunications infrastructures represent the overwhelming majority of initiators of cascading effects between critical infrastructures (Van Eeten et al., 2011). However, empirical risk approaches like dependency risk graphs require data from previous outages to determine probabilities and negative consequences. With the cases of the Northern Marianas (Van Eeten et al., 2011) and, more recently, Tonga in 2019 (Bueger & Liebetrau, 2021; Dickey et al., 2019) and 2022 (Doherty & McClure, 2022; Duckett, 2022), there are only a few incidents of internet downtime after single SCC failures for whole territories, making it hard to develop best practices. In the exemplary case of the Northern Marianas, few emergency connections could be established via satellite phones (Brodkin, 2015). Territories characterized by even higher shares of internet-dependent economic sectors such as finance and digital services or extensively digitized public administration are exposed to higher costs in the event of an internet blackout. Regarding these sectors, specific requirements could be introduced that oblige, e.g., banks, larger online businesses, or authorities to enhance redundancy through satellite internet capacities to maintain essential services. In this light, the increasing diversification of the backbone through the ongoing installation of further SCCs and broadband satellite internet technologies can be considered a positive development for the global internet backbone resilience.

13.6 Conclusion

This paper has analyzed the vulnerability of states and overseas territories by modeling the international submarine fiber optic communication network. For the vast majority (n = 126) of the territories examined, there is only a low probability of an internet outage after SCC failure. Nevertheless, academia and governments should not dismiss this situation altogether, especially if an intentional and concerted (military or terrorist) attack on the SCC network is kept in mind. As a result, however, 43 units were identified with an increased risk of cable failure, 15 of which did not even have one sufficient SCC as redundancy. In addition, we found a positive correlation between a lower redundancy level and a low socio-economic development status (developing country or least developed country). Therefore, states and territories in the Global South are more likely to be highly vulnerable to SCC faults. At the same time, they often do not offer economic incentives to implement additional SCCs.

The present study is the first to provide redundancy analysis for the SCC backbone network that has approached a worldwide perspective since 2009 (Omer et al., 2009), thus allowing a global investigation of 169 territorial units as of mid-2020. On the one hand, we followed the approach of Omer et al. in its global scope. On the other hand, we conducted our analysis with territorial entities as units, since it is at the national level that governments can take decisions for redundancy promoting measures most effectively. Thus, we followed the majority of the literature on particular continents and world regions in their national comparisons without excluding overseas territories by considering their typical insular situation.

As internet coverage continues to be built for the current 3.8 billion people not using the internet yet, there will continue to be vulnerabilities due to a lack of redundancies in developing and least developed countries. The most vulnerable territories must be identified to minimize the likelihood of such critical telecommunication infrastructure failures on the national level. For this purpose, this paper offers an initial approach based on redundancy analysis. In future research, we will pursue the inclusion of measures to assess the dependency on internet connectivity of the society and economy of a territory. Also, investigating the future impact of emerging internet-providing technologies like low-earth-orbit satellite internet and their adoption in contexts that we rated as vulnerable in this study might prove important. The satellite mega-constellations could reach the threshold for providing sufficient broadband connectivity in some time without requiring the construction of fiber optic cables. From the perspective of developing countries without any backbone connection or with low redundancy levels, a crucial question will also be whether the pricing of these services will lead to a further intensification of the Global Digital Divide.

14

CYBERWEAPONS AND ARTIFICIAL INTELLIGENCE – IMPACT, INFLUENCE AND THE CHALLENGES FOR ARMS CONTROL

ABSTRACT As cyberweapons and artificial intelligence technologies share the same technological foundation of bits and bytes, there is a strong trend of connecting both, thus addressing the imminent challenge of cyberweapons of processing, filtering and aggregating huge amounts of digital data in real time into decisions and actions. This chapter will analyze this development and highlight the increasing tendency towards AI enabled autonomous decisions in defensive as well as offensive cyberweapons, the arising additional challenges for attributing cyberattacks and the problems for developing arms control measures for this "technology fusion". However, the article also ventures an outlook how AI methods can help to mitigate these challenges if applied for arms control measures itself.

ORIGINAL PUBLICATION *Reinhold, T., & Reuter, C. (2022, October 9). Cyber Weapons and Artificial Intelligence: Impact, Influence and the Challenges for Arms Control. In T. Reinhold & N. Schörnig (Eds.),* Armament, Arms Control and Artificial Intelligence: The Janus-faced Nature of Machine Learning in the Military Realm *(pp. 145–158). Springer International Publishing. https://doi.org/10.1007/978-3-031-11043-6_11*

14.1 INTRODUCTION

The idea of the weaponization of cyber tools has been under discussion for some time (Reinhold & Reuter, 2019b; Werkner & Schörnig, 2019). Many military or national security doctrines worldwide have adapted to the development that software can be designed, injected, triggered and controlled in foreign IT systems to perform tasks ranging from espionage to sabotage. This has been done from the perspective of necessary and appropriate defensive measures but also partly as a new category for offensive planning. Although no common international understanding has yet been reached on the threats posed by cyberweapons and their prevention, let alone a binding legal instrument, this field is already beginning to change due to the emergence of improved algorithms in artificial intelligence and machine learning (AI/ML) and their potential application for or against cyberweapons (Schörnig, 2018; US-DOD, 2018c). Given the fact that cyber and AI/ML measures are natural siblings from a technical perspective, the following text provides an assessment of how AI/ML methods could influence the development of malicious cyber activities based on an overview of their current state. Regarding the threats posed by this development for international security and new challenges for arms control, the text seeks on the one hand to assess how arms control approaches should prepare for AI/ML-driven cyberweapons. On the other hand, the text also examines

the question whether and how this technology can improve arms control approaches combating the weaponization of cyberspace.

14.2 Cyberweapons and the Militarization of Cyberspace

Technological and scientific advances, especially the rapid evolution of information technology (IT), play a crucial role in questions of peace and security (Reuter, 2019). First and foremost, the most significant impact of the discussions and developments regarding the weaponization of cyberspace in recent years has been on its influence and the changes it has introduced to national and international security doctrines. An important incident has been the discovery of Stuxnet (Langner, 2013), malware developed by the US and Israel (Nakashima & Warrick, 2012) and targeted against a specific nuclear enrichment facility in Iran. Stuxnet manipulated the industrial control system of the facility by covertly changing thresholds and parameters of the control software to sabotage the enrichment process. This highly specified and hand-crafted attack on IT systems forced state leaders and decision-makers to recognize the vulnerabilities in computer systems and the threat that arises from the high degree of dependency on IT in economic, societal and government sectors. Especially critical infrastructures are now perceived to be high-risk targets for state and non-state cyberattacks. Although this was not the first cyber incident, and was hardly news for IT security specialists, the Stuxnet event demonstrated the technological possibility of crossing the cyber-physical barrier with dedicated malware and showed how to carry out actual physical destruction (McDonald et al., 2013) by remotely accessing and altering software. It also revealed the intent and the capacities of certain nation-states to develop and deploy such measures. In recent years states have reacted to this development by developing defensive measures to protect national IT infrastructures, extending national security and military doctrines to provide legal and organizational frameworks and establishing new and dedicated government or military institutions for these tasks. In addition, a large number of countries have also adopted offensive strategies, included those involving cyberspace, in their military planning and have established human and technological capacities (UNIDIR, 2013). This situation was emphasized by similar announcements by different states such as the US (US-DOD, 2018b) and the United Kingdom (Government, 2016). In 2016, NATO also declared (NATO, 2016) that incidents involving matters of or in cyberspace could invoke application of Article 5 of the Washington Treaty and prompted its member states to establish necessary military cyber capacities able to defend the alliance in this domain. A further major development was the US adoption of a new defend forward cybersecurity strategy in 2018 (US-DOD, 2018b). Declaring the ineffectiveness of defending the national IT systems by establishing IT security measures for them, the new strategy shifts activities outward to focus on the IT systems of potential adversaries and establishes a persistent engagement of cyber forces. Constant activities within foreign IT systems should, according to the strategy, provide early warning of looming attacks and keep foreign cyber forces busy enough to prevent and deter cyberattacks in the first place (Healey, 2019).

14.2.1 *The Current Situation of State-Driven Cyberattacks*

When it comes to the application of cyber measures in actual physical warfare, however, it seems that cyberattacks more often play a supporting role in military conflicts and are currently not used for massive destruction but rather for reconnaissance as well as the gathering of combat-relevant information. Most of the known cyber incidents were either cases of espionage, campaigns for political influence (Desouza et al., 2020), targeted minor IT systems or were performed with valid user credentials for critical IT systems gathered via social engineering and classic intelligence work. Although the potential for massive destruction was suspected in some cases, only a few cases with explicitly designed and deployed destructive cyberweapons have been identified so far, such as Shamoon (Tarakanov, Dimitry, 2022) or TRITON (S. Miller et al., 2019), both of which were deployed to sabotage central IT systems of Saudi Arabian petrochemical companies. From a strategic perspective, malicious cyber tools seem to have become widely accepted as an additional measure in hybrid conflicts or similar situations that deliberately stay below the threshold of full-fledged military confrontation. The relatively inexpensive creation of offensive cyber capacities – compared with traditional armament – also empowers new international actors. For instance, the Democratic People's Republic of Korea (North Korea) has become a relevant actor in cyberspace and has been responsible for different incidents over the last years (Ji-Young et al., 2019) such as the hacking attacks against a subsidiary of Sony, banks in Bangladesh or cryptocurrency marketplaces (US-DHS, 2020). Finally, the trend toward the stockpiling of vulnerabilities and exploits as the base material for cyberweapons raises new international threats. Undisclosed vulnerabilities in popular software not only provide possibilities for attacks by the withholding party but, conversely, leave anyone using the product vulnerable to attacks by any actor which becomes aware of the weak spot. The incidents of WannaCry (GReAT, 2017) and NotPetya (Mimoso, 2017), with their massive damage and commercial losses, are a dramatic demonstration of this. Both malware campaigns exploited a vulnerability named EternalBlue that had been harbored and stockpiled by the US National Security Agency (ESET, 2018). The examples demonstrate on the one hand that states are increasingly developing and deploying offensive cyber capabilities, although trying to avoid serious damage to human life and staying below the threshold of IHL-prohibited aggressive actions. On the other hand, military cyber units are probably training and preparing for utilization of their capabilities in the event of conflicts. In addition, relatively cheap military cyber capabilities are revealing potential regional power shifts, thus increasing the probability of their application in smaller-scale conflicts.

14.3 How the Technology of Cyberweapons and Its Application Will Evolve

A starting point for anticipating the influence and impact of AI/ML on the militarization of cyberspace, is the assessment of the possible evolvement of cyberweapons in general as well as consideration of future challenges regarding this type of technology. With the ever-growing automatization of all kinds of technological processes, IT systems are increasingly being integrated into physical systems and devices to control specific functions. Additionally, these IT systems will be further connected with each other (like

the Internet of Things) and to cyberspace in order to perform tasks remotely (B. Russell, 2020). This means that defense against cyberattacks will involve an ever-increasing range of distributed digital devices that need to be made even more resistant against malicious influence, as well as chain effects due to interconnections and dependencies. In addition, with the increasing number of devices and the data they create, process or store, the amount of information that needs to be integrated and processed to detect anomalies and malicious operations will continue to rise. The range of possible attack vectors will further grow and diversify. Given the necessity to react to attacks in (almost) real time, the required decision-making must be accelerated and information processed almost instantly. This requires decision-making based on integrated mechanisms of autonomy or the filtering and pre-processing of information to compensate for the relative slowness and limited capacities of human operators (Burton & Soare, 2019). Moreover, this kind of automatization might possibly lead to a cyber-vs-cyber situation, where attacks are directly blocked by dedicated defensive measures without human intervention. Similar early consideration of offensive operations and an automatic infection of possible targets within cyberspace by an NSA-backed program called MONSTERMIND (Zetter, 2014) were exposed by Edward Snowden in 2013. Following the US defend forward and persistent engagement strategy, which will probably soon be adopted by other states, such developments will result in a further undermining of global IT security by means of the preparatory or precautionary installation of backdoors within foreign IT systems, in order to have the option of deploying the intended payload in time. As cyberspace is, on the one hand, the domain of military activities but, on the other hand, also represents the physical space that processes the transmission of any kind of action, the IT infrastructures, being its backbone, will obviously become relevant targets themselves. Finally, as the capability already exists, it is presumably only a matter of time until cyber capacities will be used and deployed openly in fully-fledged military conflicts, since situations already exist where the IT of military systems and weapons themselves have become targets (Perkovich & Hoffman, 2019).

14.4 How Artificial Intelligence and Machine Learning Could Influence Cyberweapons

Reflecting on the possible impact of AI/ML on cyberweapons and the militarization of cyberspace, it is crucial to highlight that cyber and AI/ML measures are natural siblings. *"[AI and ML] share the idea of using computation as the language for intelligent behavior"* (our italic) (Kersting, 2018). From a purely technological perspective, AI/ML is *just* software: algorithms based on complex computer code that can be integrated into decision processes. Hence, AI/ML is developed and deployed within the same domain as cyber tools and to a considerable extent requires similar know-how in programming, code logic and software life cycle management. In order to be effective, cyber tools must keep pace with the latest technological developments, software updates and the modernization of devices. To reach this level of adaptability and extendibility they are often based on modern development frameworks with modularized, extendable and interchangeable software architecture (see, for example, the FLAME malware platform (Crysys Inc., 2012)). Such architecture provides an ideal platform for an extension with AI/ML components. Additionally, computer code offers optimal conditions for creating and facilitating training and testing environments for military AI/ML applications, as the environment can be defined and shaped in every specific

detail and according to the intended requirements. This reduces costs and the amount of research and development required. As described in the previous section, an important challenge for cyber as well as other military technologies is the growing amount of information that needs to be processed (Kersting & Meyer, 2018), in contrast to the decreasing time to react to incidents. This dilemma involves incidents within cyberspace but also situations where cyber tools facilitate the analysis of data and the processing of information in order to provide the basis for decision-making concerning physical systems such as weapons or reconnaissance systems. AI/ML algorithms, and especially modern approaches such as deep learning (Charniak, 2019), were developed specifically for cases involving processing large amounts of data, detecting patterns and filtering out relevant information from digital noise. According to Schörnig (Schörnig, 2018), the "spectrum of possible applications [of AI in the military] ranges from the analysis of trade data to uncover clues for the proliferation of weapons of mass destruction, to the identification of landmines that is boosted by AI with improved ground penetrating radars." Because of such capabilities, military AI applications are likely to be integrated into cyber tools, as these usually have to deal with a large amount of digital data in trying to detect relevant patterns.

14.4.1 *Explainability and Responsibility of AI-Enabled Cyberweapons*

An additional aspect of this development is that the automated conclusion process already mentioned and the resulting selection and decision about actions will be significantly changed when combined with AI/ML algorithms. Whereas the automatization of defensive cyber actions is hardly new, AI/ML are, in the sense of technology which produces an output for a given input without allowing reconstruction of the digital reasoning process or the line of thought of the machine or software that led to a specific decision. This creates situations in which the code produces decisions that are no longer deducible and thus prevent humans from intervening based on reasoning. When such AI/ML-enabled measures are used for offensive actions, this creates serious problems in connection with the necessary human integration and interaction (Schwarz, 2019). All these issues have already been the subject of heated debate in connection with autonomous weapon systems (AWS) regarding the responsibility and traceability of decisions (iPRAW, 2019). In order to address the problem of comprehensible AI/ML decisions, a dedicated field of research (XAI – Explainable Artificial Intelligence) (Gunning et al., 2019) is working on technical concepts that allow human operators either to follow the decisions during the reasoning process (ad-hoc XAI) or the decisions to be recapped once they are made (post-hoc XAI). So far, these approaches are mere theoretical concepts that lack general applicability and are hindered by specific technical features of machine learning such as the distributed and numerical representation of learned information ¸(Barredo Arrieta et al., 2020). Additionally, it is questionable whether ad-hoc explainability can be used meaningfully in an environment characterized by extremely short response times, as the two conditions are mutually exclusive. The speed of reaction in combination with the black-box character of such tools may possibly prevent any opportunity for double-checking of decisions by human operators or for their intervention. Even if the code itself does not pull the trigger, human operators might tend to trust the decisions or pre-decisions of machines and follow their suggestions due to a lack of alternatives, time pressure or perceived lack of human influence or oversight (Bajema, 2019). As AI/ML algorithms are trained for specific situations and

decisions before they are integrated into productive systems, the operators of the finished application might also be unlikely to know the specific details of the training data, nor have any chance to see, perceive or understand the assumptions and pre-conditions of this data. Besides, this inexplicability could lead to critical junctures in situations marked by high international tension. State actors on the brink of military conflict might lack the ability to communicate and explain automatically triggered actions or conclusions that led to their activities to other conflict parties, thus undermining a valuable measure of immediate conflict reduction. As unlikely as such a scenario currently seems, the discussion of application of AI/ML within the ongoing process of modernization of nuclear weapons arsenals (Field, 2019) is an example that highlights the consequences that are at stake (Boulanin, 2019). The application of AI/ML for militarized tools within cyberspace reveals an overall similarity to AWS. The debates on norms and limitations of the application of automated cyber tools could thus benefit from the lessons learned about the human role within the decision-making loop of technological systems and its consequences.

14.4.2 *AI and the Pitfalls of the Attribution of Cyberattacks*

The black-box character of AI/ML systems could also aggravate other features of cyberspace that are currently considered to be problematic, both in terms of the application of international humanitarian law (IHL) and of established norms of state conduct. One of these features of cyberspace concerns the attribution problem (Rid & Buchanan, 2015). Whereas the possibility of identifying attackers is essential for IHL and the states‘ right to use military force for self-defense (Grosswald, 2011), this task is complicated, time-consuming, and a forensic challenge due to the technical features of the cyberspace (Riebe et al., 2019). Digital information inherently contains a high degree of ambiguity and virtuality. Information can easily be copied, modified, or actively tailored to set false tracks. Consequently, the meaningfulness of information about cyber incidents needs to be critically evaluated to prevent false assumptions and reactions. Applying AI/ML measures to offensive operations will further reinforce this ambiguity and intensifies the problem of gaining a clear picture of what happened and identifying the actors behind it. The automatic AI/ML-driven evaluation of information about an incident inherently contains the problematic aspect of some conclusions about the origin of an attack being inadvertently misleading and the question of how to react proportionately. Such failure could be triggered either by incorrect or insufficiently trained algorithms, biased input information or by following intentionally created false trails (Herpig, 2019). Although the inner state of an AI is considered a black box, this condition is the result of the learning model and the data used to train the AI. Assuming that an attacker obtained knowledge of the model of an applied, static AI/ML and the data which had been used for its training – e.g., through leaks, reconnaissance, hacks, or insecure manufacturers’ supply chains – it would be possible to replicate such an AI itself and thus calculate the output that this AI/ML would generate for a specific input. Such knowledge could enable an attacker to tailor its attacks either to avoid detection or to generate incorrect conclusions (Apruzzese et al., 2019). Finally, the development and application of AI/ML in commercial, non-military IT systems, especially in the field of IT security and automated network security surveillance and defense, will produce spill-over effects in military applications. This development will increase acceptance of

such systems and put constant pressure on military decision-makers to deploy them to gain a supposed strategic or tactical advantage.

14.5 The Negative Impact on Arms Control of Artificial Intelligence in Cyberweapons

The developments outlined above add to the existing challenges involved in applying stabilizing measures in security policy to cyberspace, such as working toward peace-sustaining cyber armament reduction and cyber arms control measures. Firstly, a general problem of cyberspace is its virtual character (Reinhold & Reuter, 2019a). Data has neither a specific geographic location nor a physical representation. It can be reproduced seamlessly and is not limited to a specific and unchanging location but can instead be distributed across different places, such as in cloud applications. As explained above in connection with the problem of data ambiguity, integrating an AI/ML system into existing cyber measures further increases aspects of virtuality and non-tangibility and thus undermines established concepts of arms control even more than software itself already does (Reinhold & Reuter, 2019c). Besides obvious dual-use problems (Riebe & Reuter, 2019a) in practical terms the effortless duplication of digital data that concerns ready-made AI/ML applications as well as training data hinders the control of proliferation of military-grade AI/ML technology. This also negatively affects the ability to measure specific aspects of a regulated item, which is a core requirement of arms control (Burgers & Robinson, 2018). Like cyber tools in general, AI/ML algorithms are computer code, or even more abstractly, structured digital data. They are thus immune to any kind of countability and provide few starting points for measuring parameters that could provide meaningful classification or comparison with permissible thresholds. This missing feature also means a distinction between civil and military AI/ML systems that is capable of going beyond the mere declaration of the intended application cannot be made while also preventing any kind of classification of the capacity and performance of an AI/ML system. This situation constitutes a major obstacle to the development of viable verification approaches for AI/ML applications. Apart from that, as the performance of an AI/ML system depends to a large extent on its training, the question arises whether the trade and proliferation regulation of training data – either artificially, as tailor-made datasets or taken from real-life samples and situations – could provide a starting point for arms control and nonproliferation regimes.

14.6 How Can Artificial Intelligence Support Cyber Arms Control?

Apart from the challenges described above about how AI/ML algorithms can add to the already complicated cyberweapons debates and the attempts at peaceful development in this domain, such technologies could possibly also evolve into useful tools for cyber arms control and disarmament. In general, AI/ML algorithms are a good tool for combining and processing large amounts of different, heterogeneous, often noisy and rapidly changing data to detect patterns, regularities and hidden information (Lück, 2019). A specifically powerful aspect of this technology is the ability to identify similarities within data and find useful matching items that do not fully correspond to the trained items but relate to them with a high degree of certainty. This kind of detection quality is

usually a problem that cannot be solved with hard-coded deterministic rules. By contrast, an AI/ML algorithm is able to identify relevant detection parameters during its training phase, establishing a self-developed filter for relevant and irrelevant information. As a result, AI/ML algorithms could prove to be the right tool for managing the information overload of IT systems (Kaufhold et al., 2020) and the challenge of finding the needle in the haystack. Such challenges could be the task of searching for anomalies in information provided by states in the context of confidence-building measures or processing surveillance imagery to detect military installations. A meaningful, currently unexplored application could be to control the proliferation of cyberweapons (J. Silomon, 2018) by monitoring the distribution and occurrence of specific parts of weaponized computer code. As already mentioned, code can easily be copied and will, in almost all cases, be slightly modified or extended to fit into existing cyberweapons, to work with the specific tools and programming frameworks, or to match specific target criteria. Any detection mechanisms searching for an exact piece of computer code will presumably fail to detect such modified versions. An AI/ML algorithm could be trained to circumvent this problem and to provide at least indicators and probability measures of whether and to what extent computer code matches a specific sample. A similar approach could be used to detect and identify actors behind cyberattacks. Even if this is not directly a task of arms control, it overlaps with the regulation of cyberweapons, because an actor is visible, detectable and identifiable by its behavior, by technical operations performed in foreign IT systems and by the tools employed (Sibi Chakkaravarthy et al., 2019). Whereas it is possible and common to counterfeit these indicators in order to lay a false trail, an AI could be used to detect unconscious similarities of the attackers' style, habits and methods. Institutionalized military cyber actors in particular develop their know-how and the required skills over time. They create, extend and modify their own tool sets and cyberweapon arsenals, which are then reconfigured, combined and adjusted for a specific operation (Olszewski, 2018). This means that specific actors often have digital fingerprints regarding their customary tools and hacking strategies. Nearly every cyber activity creates digital traces such as small pieces of code that attackers have previously used to perform their tasks, manipulate files, change system settings or log entries or IP addresses of remote IT systems where data has been copied. Such detectable traces are called samples and are already used to compare new code to known samples from prior incidents in order to draw conclusions about an alleged actor. Although captured samples like these rarely match existing samples perfectly, they do contain similarities as they come from the same complex cyberweapon project, use similar methods and approaches, or are more advanced versions of each other. Detecting these similarities and identifying cyberweapons is a task where AI/ML approaches and algorithms are highly suitable (Roberts, 2019). For example, such identification measures are already used by IT security forensics when analyzing cyber incidents (Känzig et al., 2019). They are often combined with further indicators such as specific habits and ways of programming, the structuring of computer code or recurring phrases and names. Lastly, the black-box character of AI/ML applications could also be an advantage for arms control measures. An essential element of practical control and compliance monitoring of arms control regimes is the requirement that the actors involved do not want to disclose any sensitive information about the regulated or controlled item (Kütt et al., 2018). This requires technical procedures where participating parties – usually states – are required to disclose as little information as possible when verification is performed and verification devices are developed that conceal all processing steps. In addition, the participating parties would have to be convinced that the results will be reliable and trustworthy. Such a tool, in which a defined input leads to a binary decision of is or is not

a weapon, could be achieved through AI/ML procedures. To prevent doubts regarding the reliability and the acceptability of the algorithm's decision it would be necessary to prevent any modification or tampering and to preserve the integrity of the algorithm and its trained state. This could be achieved by securing the AI/ML application with digital seals, cryptographically calculated unique values – usually very long numbers – like checksums and hashes that represent a specific state of arbitrary digital information. A recalculation of the digital seal would immediately reveal any modification as it would result in a different number if the information has been changed (Putz et al., 2019). These mere outlines of applicable approaches presumably have other peculiarities that need to be taken into account when it comes to real-world applications. Although this issue goes beyond the scope of this chapter, it shows that, despite new challenges, AI/ML approaches can also contribute to arms control.

14.7 Conclusion

This assessment has provided an overview of the possible development and impact of AI/ML methods on cyberweapons. It is based on current trends and technical AI/ML developments as well as on the already ongoing application of or research on AI/ML in other military fields of operation. The assessment shows that the military application of AI/ML for cyber related tasks will probably exacerbate an already tense situation involving a cyber arms race on the one hand and a lack of international measures to prevent destabilizing and harmful effects on the other. Established measures for arms control, whose application to cyberweapons is already hindered by specific technical features of these tools, will face further challenges. Furthermore, for military decision-makers AI/ML algorithms seem to provide solutions for enhancing their weapon systems and battlefield management capabilities through their ability to integrate, process and refine large amounts of digital data. This could provide a strong incentive for military decision-makers to pursue and apply these approaches. However, the assessment also showed that, in addition to the necessary questions of peace and conflict research regarding AI/ML in cyberweapons, technological developments reflect ongoing debates about lethal autonomous weapon systems. This makes it possible to participate in these discussions and to benefit from lessons learned. Finally, AI/ML approaches could also provide valuable insights into the challenges of arms control for cyberweapons and help to circumvent some of its technological pitfalls. Either way, artificial intelligence and machine learning are just beginning to find their way into military cyber systems, and the time has come to critically accompany this trend and conduct further research in order to promote peaceful development of cyberspace.

PUBLICATIONS
PART C

15

PREVENTING THE ESCALATION OF CYBERCONFLICTS: TOWARDS AN APPROACH TO PLAUSIBLY ASSURE THE NON-INVOLVEMENT IN A CYBERATTACK

ABSTRACT While cyberspace has evolved into a commonly shared space vital to our individual lives and societies, malicious cyber activities by state actors as part of espionage operations, regarding defense strategies, or as part of traditional conflicts have strongly increased. In contrast, attributing the origin of such activities remains problematic. The ambiguity of digital data raises the problem of misinterpreting available information, increasing the risk of misinformed reactions and conflict escalation. In order to reduce this risk, this paper proposes a transparency system based on technologies which usually already exist for IT security measures that an accused actor in a specific incident can use to provide credible information which plausibly assures his non-involvement. The paper analyses the technical requirements, presents the technical concept based and discusses the necessary adjustments to existing IT networks for its implementation. Intended as a measure for conflict de-escalation, the paper further discusses the limitations of this approach, especially with regard to technical limits as well as the political motivation and behavior of states.

ORIGINAL PUBLICATION *Reinhold, T., & Reuter, C. (2023a). Preventing the Escalation of Cyber Conflicts: Towards an Approach To Plausibly Assure the Non-Involvement in a Cyberattack.* Zeitschrift für Friedens- und Konfliktforschung (ZeFKo). *https://doi.org/10.1007/s42597-023-00099-7*

15.1 INTRODUCTION

Disruptive cyber operations, influencing campaigns and espionage are increasingly a daily occurrence in this domain. This is underscored by the fact that states have included such operations in their domestic security and defense doctrines and understand this as an additional field of military operations (UNIDIR, 2016). Offensive activities such as cyberattacks can be launched and performed indirectly or with counterfeited digital footprints, i.e. incriminating uninvolved, innocent actors (Warrell & Foy, 2019). However, due to a complex and often time-consuming process that impedes effective counter activities, many have declared the secure identification of sources of malicious cyber activities to be impractical (UNIDIR, 2018). These technical, political and organizational challenges of the so-called attribution problem increase the risk of misunderstanding, miscalculating and misinterpreting malicious cyber activities. The precise role and influence of military cyber activities in open conflicts is still ambiguous. However, particularly the increasing attacks against critical infrastructures (Lunden et al., 2021; Noguchi & Ueda, 2017), the fear of interference by foreign states in vital public services

in peace times, or situations where immediate responses are necessary to counter and mitigate cyber threats could lead to misguided responses and create a momentum for the escalation of conflicts. Although mostly considered a last resort, some states have even reserved the right to respond to cyberattacks with physical military means if deemed necessary as an immediate defense against hazards. This scenario and its increasing probability have recently been expressed, e.g., by US President Biden, with regard to the ongoing cyber tension between the US and other major powers (US White House, 2021). These problems may be exacerbated if other ways of crisis communication or security and trust-building measures between adversaries are missing. Whereas the history of technical peace research has addressed these issues for other military technologies with conflict prevention and de-escalation measures, suitable approaches for peacekeeping in cyberspace are currently lacking and their development is strongly recommended (Wissenschaftsrat Deutschlands, 2019). Against this background, this paper focuses on the research question of how technical measures can enable state actors to mutually control their cyberspace activities of military forces and intelligence services by providing verifiable data that can be used to assess and verify a state's non-involvement in a specific previous or ongoing cyber incident. This paper does not aim to provide a ready-made implementable measure but rather to provide a food-for-thought as well as a technical foundation for such an approach and to examine the possibilities for establishing such measures based on existing IT infrastructures. This paper further discusses the necessary technical adjustments required to provide this kind of information while maintaining an appropriate level of secrecy. Given the sensitive nature of internet traffic surveillance, the paper discusses the limits of such a measure with regard to its implementation in order to uphold human rights principles and avoid their violation. Concerning the aspect that cyberattacks are often carried out by hacker groups or other so-called proxies that are not directly associated with or under the control of state institutions, the paper also analyses the political preconditions and limitations. Regardless of the quite specific and small application scope and its requirements and restrictions, we hope to provide an impulse for cyberconflict de-escalation measures that help to mitigate the escalation risks inherent to the status quo. Finally, given the relatively new domain of cyberspace, the paper aims to connect computer science with peace and conflict research and hopes to provide valuable impulses for dialogue between them (Reuter, 2019).

This paper is structured into the following sections: After the introduction and the definition of the research question in the first section, current research on the attribution problem based on related work from a technical as well as political perspective is discussed. Section 15.3 presents the cases that have been selected to illustrate this work's background and motivation. Section 15.4 explains the required technical principles of network-based digital data transfer and the problem of the ambiguity of digital data. Afterward, section 15.5 discusses the motivation of states to join a risk reduction measure and comprises the central arguments of this paper with a conceptual as well as a technical outline for a system that can provide evidence to verify an accused state's non-involvement in a cyberattack. Finally, in section 15.6, the developed measures, their limitation, and potential pitfalls are discussed in relation to the research question and their practical application. The section ends with an outlook on further extensions of this approach as well as future research opportunities.

15.2 Related work

The problem of attributing malicious cyber activities to their perpetrators has gained much attention in recent years. As attribution is both a technical and a political process, many different approaches have been proposed to solve the so-called attribution problem.

15.2.1 *Technical challenges*

A common argument for most approaches is that attribution is a complicated task, as information needs to be collected and interpreted forensically (Koch & Golling, 2019). To support and improve this time and resource-consuming process, one area of research focuses on a more detailed and standardized implementation of data collection and storage based on frameworks (Lilly et al., 2019). This is also partly linked to advanced intrusion detection systems (Rubio et al., 2019), which are able to detect attack attempts for early data acquisition of the attacker's activities (Ni et al., 2016). Other investigations focus on the topology of IT networks and the question of where intrusion detection systems should best be placed, and how the processing of physical signals alone can indicate uncommon behavior and possible attacks (Giraldo et al., 2019), sometimes alongside methods of artificial intelligence and machine learning to detect intrusions and to identify the attacker's location (R. S. S. Kumar et al., 2017). The approach aims to create complex data sets by including threat intelligence sharing platforms to gain a broader information basis (Perry et al., 2019). This allows for classifying and filtering the aggregated information on cyber incidents or applying data-analysis methods to detect similarities in attackers' behavior (Hoon et al., 2018; Shute et al., 2017). Other approaches focus on the inclusion of publicly available open-source information (Lemay et al., 2018) or social media (Bargar et al., 2019; S. Kumar & Carley, 2016). In summary, current research on attribution as a measure of security concentrates on the tasks of data acquisition, optimization and analysis to provide a better and faster answer to the question: *Who did it?* In contrast, the perspective of this paper focuses on the question: *How can I credibly assure that I was not involved/that I am not the perpetrator?* This aspect is essential in order to improve security by preventing misattribution. This is a novel approach that has not been scientifically researched so far but lies in the tradition and demands of technical peace and conflict research (Reuter et al., 2020).

15.2.2 *Political and security challenges*

Attribution is and "should not be an aim in itself" (Broeders et al., 2020), as it pursues an intent and the "decision to attribute a cyber operation to another actor should be strictly linked to a broader policy objective(s) that a state or a group of states wishes to achieve" (Broeders et al., 2020). This broader objective has been discussed in different ways. First, attributing an attack towards a designated attacking country is a prerequisite for any legal military response in accordance with the UN Charter (Wingfield & Wingo, 2021). In other words, without sufficient evidence, there is no target to refer to legitimately. Rowe (Rowe, 2015) focuses on this aspect and further discusses (1) possible measures to achieve attribution, (2) the complexity of this task from a technical and legal standpoint,

and (3) the influence this can have on an attacked state's political and military decision-making processes. In most cases, attribution requires cooperation between states in order to collect technical evidence of an attack, thus requiring a mutual understanding of the problem, legal common ground and suitable corresponding processes (UNGGE, 2015b) for the collection as well as the exchange of threat intelligence (Riebe et al., 2019). This is discussed by Bendiek and Schulze (Bendiek & Schulze, 2021), with a focus on the development and enforcement of a harmonized cyber sanction regime of EU member states. A further aspect of attribution is discussed regarding its applicability and limitations for deterrence and how "scaling of exploitation and retaliation costs lead to different degrees of coverage and effectiveness for deterrence by denial and punishment" (Lindsay, 2015). A related aspect considers the role of public attribution – naming and shaming of the accused country – and the trust and accountability problem it faces regarding national interests (Egloff & Wenger, 2019). A prominent approach to overcome this problem whilst fostering transparency in attribution processes is the establishment and empowerment of independent, supranational institutions that could be in charge of attribution based on evidence provided by attacked countries (Davis II et al., 2017; Droz & Stauffacher, 2018). Other debates address the role of non-state actors behind cyberattacks, the responsibility and due diligence of states, and how attribution and the right to self-defense apply under these conditions (Starski, 2015). Other research explores the question of whether or to what extent military cyber activities have shaped the battlefield and how this new military domain influences strategic or tactical military approaches in open conflicts (Kostyuk & Zhukov, 2019). Blagden (Blagden, 2020) even argues that the technical challenges of revealing the concealment of attackers can be neglected, as an "attacker must necessarily reveal a set of interests that it values", putting the question of interests at the center and declaring attribution a primarily political task.

15.3 Case examples

The research question is motivated by real-world cyber incidents and the associated escalation risks. In order to illustrate them and the constraints of a measure focused on state actors, we have selected two examples of actual critical cyber-based conflicts that happened in the context of strong regional political and military tensions between actors that have developed strong offensive cyber capabilities over the last years. The examples are used to derive an exemplary model for the attribution process and to define requirements for a technical system that can provide plausible information that provides evidence and assurance for non-involvement in a specific cyber incident. The findings are put together into a model for a technical measure and its applicability. The security gain, as well as its limitations and necessary further extensions, are discussed both for the technical and the political context.

15.3.1 *Selected case examples*

The following two brief examples illustrate the different scenarios and variations of cyber-based incidents in the context of regional tensions and the problems of unclear or misleading attribution. They have been selected because they prototypically exemplify the ambiguity of digital data as well as the situational pressure for action, whereas the

overall political situation suggests an obvious answer to the question of the origin of the attacks, thus increasing the risk of misinterpretation and the danger of misreaction. Sources about such incidents, the political considerations regarding countermeasures and the actual measures taken are often rare or incomplete – partly because the public (non-)communication towards the suspected attackers is often already part of the reaction. In most cases, the only available sources are press reports and other anecdotal public media coverage. Compared to other incidents, the selected case studies are based on credible sources that at least provide enough details to analyze the course of events.

The cyberattack against chemical plants in Saudi Arabia, August 2017

In August 2017, a cyberattack was detected at a petrochemical plant in Saudi Arabia that targeted the industrial control systems which monitor, control and regulate all different aspects of the industrial process (Perlroth & Krauss, 2018). In contrast to former attacks against such systems, which have often tried to silently manipulate the controlled processes, the aim of the detected attack was presumably to deliberately sabotage the industrial hardware by triggering explosions. This would most likely have resulted in significant damage to the facility as well as possibly human injuries or casualties. This was prevented by a programming mistake made by the attackers. Investigators blamed Iran for the attack due to previous incidents against governmental institutions and industrial facilities (Tarakanov, Dimitry, 2022) as well as overall political tensions between the two countries (Baezner, 2019; Marcus, 2019). However, cyber capabilities from other opposing nations, such as China (The State Council Information Office of the People's Republic of China, 2019), Russia (Karaganov & Suslov, 2019), or Israel (Dwyer & Silomon, 2019) would also have been sufficient to conduct this attack. Press reports suggested that Saudi Arabia feared an immediate second attempt of sabotage after the failed first attempt. This required a quick decision on whether and how to respond. As official communication channels between the nations have been scarce since an attack on the Saudi Embassy in Tehran in 2016 (Gazette, 2016), the situation highlights the dangerous scenario of the lack of significant evidence and political crisis communication channels.

The cyberattack against the Ukrainian power grid, December 2015

The second case is the cyberattack against five power supply companies in Ukraine that took place on December 23, 2015 (Zetter, 2016). The attack itself targeted the control systems of the power plants and their supply infrastructure, as well as the companies' call-center services so that customers could no longer receive any information. In total, up to 230,000 people were cut off from electricity for one to six hours. The attack occurred in the aftermath of the Russian occupation of Crimea as part of the ongoing crisis between Ukraine and the Russian Federation. The attack was immediately attributed to hackers from Russia based on the geolocation of the attackers' IP addresses. From a technical perspective, this is not itself valid or sufficient evidence and could have been a false track laid by third-party attackers, demonstrating the ambiguity of the available information. So far, it is not entirely clear if the cyberattack was a sabotage operation, a test run of its own capabilities, or a political demonstration of power. In any

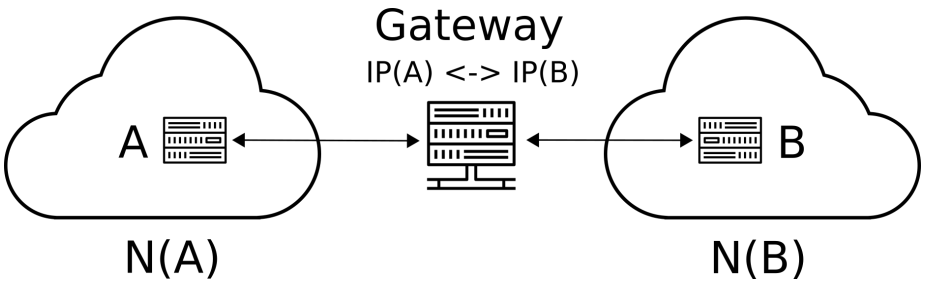

Figure 15.1: Simplified model of data transfer between two computers in separate networks (Source: own illustration)

case, it was an attack on critical infrastructure and, thus, a violation of established rules of international humanitarian law (IHL).

15.3.2 *Comparison of the example cases*

The selected cases, with their different presumable motivation of sabotage and social disruption and insecurity, have in common that the involved actors are in a situation of high political tension and ongoing small-scale military conflicts or are on the brink of tensions erupting. Although the disputes have so far remained below the threshold of a declared war, tensions have increased due to the decline in communication relations and the loss of commonly shared forums as well as diverging interests. The available sources do not allow for any conclusions as to whether a technical incident analysis had been performed. Given the presented expense for a valid attribution and the short time available in all cases, it is highly unlikely that forensically robust results had been available. As a result, the attribution of cyber incidents was strongly influenced by the overall situation and previous incidents. The probability of blaming the usual suspect when other information is missing or incomplete is very high in such cases, increasing the risk of misattribution and escalation of the incidents into a full-fledged military confrontation even further.

15.4 Technical Properties of the Cyberspace and the Ambiguity of Digital Data

The difficulties related to attribution in cyberspace and the risk of false attribution due to misinterpretation of the available information are based on some specific technical features of this domain and the way in which cyberattacks are performed. The following section will highlight these particularities as a prerequisite for the research question.

Cyberspace is a virtual domain by design that abstracts a space from a specific real existing geographic location. It consists of autonomous, self-contained networks that integrate and connect groups of different IT systems or sub-networks. The networks are connected via gateway servers at their point of contact (borders). To perform any kind of data transmission between two IT systems, its necessary to identify both systems,

which is done by using a technique named Internet protocol addressing or, in short, IP addressing (Scaglia, 2007). Figure 15.1 presents a simplified model of such a data transfer and the involved IT systems. It is important to understand that an address of an IT system – in the following called (A) – is not inevitably unique. It must only be distinct within the network to which the system is directly connected, hereafter referred to as N(A). Any connection of (A) to an external IT system (B) that is not part of N(A) is transferred over the gateway server that – in the simplest case – directly connects the networks N(A) and N(B) but in actual scenarios involves multiple networks. The gateway server handles the necessary network address translation (NAT) to ensure that data can be transferred between two networks with possibly non-unique IP addresses (Juniper, 2022). This means that the effective sender address that (B) can identify while receiving data from (A) is not a unique address of this IT system but rather an address provided by the gateway server of N(A), which means that there is no clear and directly visible and re-traceable path to the origin of the connection. This aspect also entails that any kind of geographical localizing based on IP addresses will reveal the involved networks and not necessarily the specific IT system (A) itself. Over the last years, the current protocol – the technical rules of how network traffic is handled by the different IT devices – has been shifting towards a more modern approach (Internet Protocol version 6 or short IPv6 instead of IPv4) where devices have worldwide unique IP addresses (Juniper, 2022). Nevertheless, for reasons of data protection or security, these unique addresses are often reduced to their network part, and the identification of a single IT system is taken over by gateway servers. Another significant aspect of the technological basics of cyberspace is that it abstracts the process of data transmission between IT systems over different structural and conventional layers and generalizes specific functionalities with common technical protocols (Scaglia, 2007). All IT systems that communicate over cyberspace have to use these protocols. There is no close connection between the observed usage of an IT system – like a cyberattack – and its real-world and intended purpose. Even a forensically waterproof identification of an attack's origin cannot exclude the possibility that an identified IT system had been taken over by adversaries. A popular but mere theoretical but conceptually valid example is the misuse of IT systems in a hospital that had been hacked to carry out cyberattacks. A third important principle of cyberspace concerns the aspect that any data transmitted during connections between two distant IT systems (A) and (B) is split up into a large number of small packets that are sent separately and merged at their destination. Each transmitted packet can potentially take a different path, i.e., route. This principle guarantees that disruptions of a route can be balanced by other transmission paths. In the context of cyberattacks, this means that retracing the steps of attacks to their origin equals finding the path back via potentially different and numerous routes. Drawing from these specific technical features, many real-world cyberattack scenarios involve multiple steps of intermediary hubs that are used to blur tracks. This often involves the usage of one or more so-called command and control servers (C&C) that are used by attackers to coordinate progress and collect stolen data. C&Cs are either hijacked systems or rented servers that do not belong to the attackers themselves. Considering the explanation of IP addresses provided above, this means that even if a victim of a cyberattack can identify a unique IT system via its IP address, it is probably not the actual system of the attacker. Therefore, the task of attributing such attacks typically involves the analysis of at least some of the IT systems used as hubs and the C&C infrastructure. Aside from the time required to perform these actions, each step potentially relies on the cooperation of other actors to gather information from affected systems within their jurisdiction, as well as the availability of such data samples (Clark & Landau, 2010).

These discussed features of cyberspace complicate the attribution and create a strong character of ambiguity. Available information on attacks and traces towards attackers is, in most real-world cases, either incomplete or inconclusive in terms of its interpretation. In addition, this information is easy to manipulate or counterfeit, and attackers might have created false tracks by forging misleading evidence, commonly described as false flag operations (Steffens, 2020). On the other hand, cyberattacks against critical systems often require immediate decisions to be made about countermeasures to stop the attack's ongoing threat. In combination, the current lack of internationally binding norms for responsible state behavior in cyberspace leads to the situations described, in which misunderstandings, miscalculations, and misinterpretations could cause wrong and potentially destructive responses.

15.5 Reducing the Escalation Risk: Outline of a System to Plausibly Assure Non-involvement in a Cyberattack

Based on the previous assessment, this section proposes a concept that, while it cannot help to diminish the *burden of proof* of the cyberattack victim, aims to help to reduce the threat of accidental escalation of a conflict. The concept is understood in the sense of the CSCE Helsinki final act that recognized "[. . .] the need to contribute to reducing the dangers of armed conflict and of misunderstanding or miscalculation of military activities which could give rise to apprehension, particularly in a situation where the participating States lack clear and timely information about the nature of such activities" (CSCE, 1975). The measures are based on the idea that a reduction in the escalation risk or even its prevention can be achieved if an accused actor is able to credibly and plausibly assure they were not involved in a specific cyber incident by providing verifiable information. Although it is ultimately the decision of an attacked state how to react, which is often determined by various aspects, political goals or political signals, such information could alter its judgment and assessment of the situation. By being as transparent as deemed necessary about current cyber activities within their military or intelligence services networks, accused states can further signal to other states that the risk of an imminent cyber threat is unlikely to exist.

15.5.1 *Political incentives and motivation of states to establish and comply with such a measure*

States as actors often have diverging interests and are – besides treaties or other binding commitments – able and sometimes willing to act in contradiction (e.g., by cheating) to their commitments. The resulting limitations will be analyzed and discussed in more detail in Chapter 15.6.1 Regarding this context, the main motivation of a state to establish measures that can plausibly assure their non-involvement in a specific cyber threat is their self-interest in preventing the escalation of a conflict or of being falsely held accountable for malicious activities. Being able to provide the described information and to make their own cyber activities transparent could even be a measure of confidence-building. In addition, the system will be completely operated by and, therefore, under the control of the establishing state or authorities that established the measure, and – in the case of an incident – only the state decides which information

is disclosed. All this, of course, does not diminish the possibilities of carrying out malicious operations covertly nonetheless or of even being the actual perpetrator of the cyberattack in question (e.g., by using a proxy). The core element, in terms of the measure's political effects, is the credibility of an accused state. The de-escalation effect of the measure is directly linked to this credibility and, therefore, its other military or intelligence cyberspace activities and – in the best case – its refrain from using covert or proxy operations. In conclusion, this means that it is in a state's interest to create opportunities to de-escalate and avoid cyberconflicts and to maintain its credibility.

15.5.2 *Requirements for a system for the plausible assurance of non-involvement*

In order to assure non-involvement in a cyber incident, an actor – besides their political credibility – needs to supply verifiable data that fulfill the following requirements:

- *Relevance*: The information must contain all incoming and outgoing relevant network connections of the accused actor: (a) to or from all networks of the attacked actor, (b) to or from the IT systems that had been targeted and (c) about connections to networks or IT systems of third parties that are suspected of having been used as C&C infrastructure or any other kind of indirect attack controlling measures.
- *Sufficiency*: The above information must be supplied for the incoming and outgoing connections of a defined scale of networks that are under the accused actor's jurisdiction to a technical degree that allows plausible verification and ensures non-involvement.
- *Timeliness*: The information must be supplied in a timely manner for the entire period of the cyberattacks and/or the malicious activities.

The information could either be supplied voluntarily by an actor or as a response to a request by an accusing actor or entrusted authority. The provided information can be anonymized to the degree that assures non-involvement in a specific attack while filtering out irrelevant data or disguising secret information. This is possible, e.g., by reducing the logged number of connections to such an extent that the subnets can be identified, but not the actual machine. For IPv4 as well as the newer protocol IPv6, this can be done by cutting the IP address of each logged system, IP (A/B/...), down to a subnet address. This would successfully hide the actual amount of different IT systems within this subnet as well as the individual connection information of each of these IT systems. On the other hand, it allows the mapping and comparison of all outgoing and incoming connections of this network that could be associated with a cyberattack that occurred at the same time, based on the logged timestamps, destinations and type of traffic. Based on this information, either the accusing actor or a neutral third party would be able to assess the provided data. Instead of tracing back the path to the alleged attacker, the validation would be able to directly focus on the supposed origin of the attack path and therefore be able to validate a statement of non-involvement. As has already been pointed out, this will not reveal the identity of the actual attacker but can help to relieve the alleged attacker.

15.5.3 *A conceptual cyber incident model as a potential use case scenario*

Before discussing the technical implementation, the following section presents a schematic model of a cyber incident, as illustrated in Figure 15.2 It is set in a crisis situation between two previously introduced actors – (A) and (B) – with an assumed strong degree of mistrust, negative expectations, and non-communication, as well as current political or military tensions. In this situation, actor (B) was the victim of a cyberattack. The example further involves the actual attacker actor (C) and the uninvolved but exploited actor (D).

1. Actor (B) detects an incident, in the following referred to as (x), within its networks N(B).

2. Entitled authorities of Actor (B) check the logged information as well as the technical integrity of the affected IT systems and detect unauthorized access to these systems from a source outside of N(B) over a time frame called T(x).

3. Actor (B) identifies the unauthorized access from an IP address IP(x) that is registered to a party within the jurisdiction of actor (D). Actor (D) is assumed to be uninvolved in the current conflict and has no history of aggressive behavior against actor (B).

4. Due to specific circumstances, the authorized agencies of actor (B) are not able to trace back the path from (x). Possible reasons for this situation, as discussed before, are:

    - The short reaction time that is available to decide on countermeasures by actor (B).
    - A refusal of actor (D) to provide further information.
    - The absence of valid logging information either on the identified systems of actor (D) or on further intermediary steps.

5. Actor (B) accuses actor (A) of being the agent behind the incident (x) with reference to the political background, ongoing tensions, former incidents, or aggressive announcements by actor (A), or due to similar false-flag operations that had been tied to actor (A) in the past. To bring the harmful cyber activities to an end, actor (B) signals the willingness to use strong political or economic measures (the most likely reaction in espionage scenarios) or military force (the likely reaction in cyberattack scenarios).

In terms of this described scenario, the rephrased research question is: By which technical measures can Actor (A) credibly and plausibly assure that no IT systems within N(A) had any connection to the identified attacking system IP(x) for the time frame T(x) of the incident?

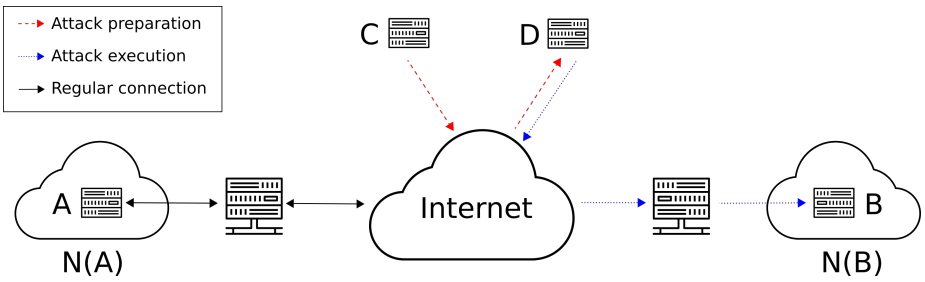

Figure 15.2: Schematic model of common cyberattacks via an intermediary third party (Source: own illustration)

15.5.4 *Necessary technical capabilities and specifications*

The credibility of the provided information and thus the plausibility of the argumentation of a states own non-involvement rests upon two technical properties of the data: (a) the tamper-proof collection and storage, and (b) the level of detail, the coverage and the verifiability of the information collected and provided by an accused actor. Therefore, based on the conceptual assessment of section 15.5.2, this data and its collection measures need to fulfill the following technical specifications:

- Cover a time frame that is long enough to satisfy the accusing actor.
- Collect data from all relevant IT networks without hindering their functionality.
- Contain details on the endpoints of all connections that had been established and should rely on information that is always accessible during the network-based data transmission.
- Use a tamper-proof, non-circumstantial data acquisition and storage measure that is considered trustworthy even in non-cooperative actor relationships.

Notably, all IT network technologies gather and potentially store information on the established or performed network connections. This information analysis and storage already takes place in most cases as an IT security measure to monitor connections, identify malicious activities, detect or track hacking attacks and attempts(Ekran Systems, 2022; Gür et al., 2015), and control the access to IT systems and networks (so-called firewalling, and the measures we propose mostly require no changes apart from immutable storage. Under these prerequisites, the capability to gather the necessary information is taken for granted and will not be described any further. In terms of the proposed context as a risk-reduction measure and with regard to the research question, this focus emerges into the following aspects:

- In which IT networks and where within a network does the data need to be collected?
- What kind of data needs to be stored and to which level of detail?

- For how long should the data be stored?
- How can tamper-proof storage be performed?

These questions will be discussed in detail in the following sections. Moreover, chapter 15.6 will discuss the limitation of this approach and the possibility of state cheating.

In which IT networks and where within an IT network does data have to be gathered?

Connected IT systems are always topologically organized in network structures on the physical level, where the gateway servers between networks know about any outgoing and incoming connections for a specific network. With regard to the network-sub-network topology, it is only necessary to store the information about network activity at the logically outermost gateways, where data leaves the IT system of an actor and is transferred to external systems. In terms of the context of an application, only networks of institutions under direct state legislation or control and likely to be responsible for cyber activities in foreign IT networks, such as military networks or the IT systems and networks of intelligence services, need to implement the measure. While this has its limitations (see Chapter 15.6 for further discussion), it prevents the establishment of unlawful surveillance measures. Therefore, this storage needs to be performed on all gateways that connect these specific networks to the outside world – like private or commercial networks – to prevent hidden channels. The measure itself may need additional capacities for data storage but does not affect the functionality of the gateways nor the question of who runs them. As argued, in most cases, the relevant information has already been gathered and is ready to use.

What kind of data must be gathered and stored, and to which level of detail?

A typical connection between two IT systems consists of the exchange of multiple data packets with different purposes that establish the connection, transfer the data in multiple single packets, acknowledge the successful transmission of the packets, and finally close the connection. Besides the actual payload, each packet contains the so-called metadata of the information on the data packet sender and its destination – both identified via IP addresses. In terms of the proposed measure, the following information needs to be stored to ensure that for a given time slot, no data was transmitted to a specific IT system or network:

- The timestamp of when connections were established and closed, either from within the network or by request from external IT systems.
- The destinations (for outgoing connections) or origins (for incoming connections) of connections.
- The amount of data that has been transferred and, if available, for which kind of application the transferred data is meant to be processed.

For an effective application of the proposed measure, it is therefore not necessary to store the transferred packets but only the mentioned metadata on the connections. These kinds of data are simple text-based information that can be stored and transmitted without difficulty.[1]. Although the actual logging rules for each implementing system are highly system-dependent, the above information should be gathered and stored for each connection that passes through the system without exception. In terms of secrecy, the stored information could potentially reveal sensitive data, as it contains details on the quantity and types of IT systems and services within the network, as well as the quantity and locations of systems that the specific gateway usually connects to. In order to maintain confidentiality, IP addresses can be partly anonymized to contain only information on the sender and destination networks, as it still contains sufficient information to assure non-involvement. This aspect also addresses upcoming IT security measures of moving target defense (MTD) (Carvalho & Ford, 2014; Dishington et al., 2019), which aim to obfuscate the identity of IT systems within a network to confuse attackers by, for instance, randomly changing IP addresses.

How long does data need to be stored?

The question of the storage duration cannot be answered uniformly and is rather a task of considering a trade-off between essential storage resources, secrecy, and evidential value; notably, this parameter is easily adjustable. A solid basis for an estimated storage time can be provided by studies that are regularly performed by IT security companies that analyze hacking incidents. For example, a report by Mandiant Consulting (Hau et al., 2016) estimated that in 2016, cyberattacks had lasted 146 days on the worldwide average before they were detected (Hau et al., 2016). The same report calculated the average detection life span of hacking attacks for Europe and the Middle East to be up to 469 days. The analysts calculated a decreasing worldwide average of 99 days for 2017 and 56 days for 2019 and concluded that the life span of attacks significantly dropped due to higher sensitivity for IT security (Mandiant Corporation, 2017, 2020). Another approach to further specify the necessary logging time frame could be taken from recommendations (IBM, 2021) for the size and time frame of logging data structures for IT security reasons. This system-specific assessment is performed to determine how far back in time a hacked target can retrace steps within its own systems via logged digital forensic information. Further inferences can be drawn from national and international data retention policies to track criminal cyber activities (E.G.N.Y.T.E., 2021). Furthermore, in the case of an ongoing attack, providing the current gateway activities of an accused actor can be an important measure to provide evidence for their non-involvement in the communication of the actual attacker with their command-and-control infrastructures. In the optimistic circumstances that our proposed measure is established as part of a treaty, the logging time frame should be commonly agreed upon and defined for all treaty members, as any past activity that is older than the time frame will not be reflected in the logs.

[1]For illustration, storing such simple text-based information, one gigabyte of storage can hold up to 678,000 pages of text. (https://www.digitalwarroom.com/blog/how-many-pages-in-a-gigabyte). The King James Bible (Old and New Testaments), which is stored at Project Gutenberg in plaintext ASCII format (7 or 8 bytes per character) has a size if around 4.2 Megabytes.

How can the process of data gathering and storage be technically tamper-proof?

The acquisition and storage of logging information is not new and is a common feature of IT security measures. In the context of this proposal, it is the credibility of such data that decides whether a targeted victim believes the digital facts that an accused actor provides. Credibility can be reached by technically ensuring that neither the process of the logging data acquisition has tampered with (for example, connections to some specific endpoints get excluded from logging) nor that logged information can be manipulated afterward. Preventing and ensuring tamper-proof data storage is an issue that can be solved by cryptographically and incrementally signing the logged data, a method that is known as immutable data storage (Rovnyagin et al., 2021). This kind of technical verification for streams of logging data is a concept that has already been described as an audit log or audit trail (Schneier & Kelsey, 1998) for use cases in safety or secrecy critical scenarios (Putz et al., 2019). An additional degree of credibility can be achieved by ensuring that the mechanism which collects the logging information (commonly defined by so-called logging rules) itself has not been modified. This is possible by including the logging rules or hashes (digital fingerprints) of the logging code as part of the cryptographically secured logged data, as it provides tamper-proof copies of the logging process and its configuration for a retroactive validity comparison. Creating and securing logging data with an immutable storage mechanism results in an increase in necessary processing and storage capacities as the information has to be encrypted and digitally signed. However, recent developments show that this can be done highly efficiently by using the GPU (graphics processing unit), a commonly used component in modern computers and that the storage scales 1:1 with the stored information items without creating overhead. The actual storage capacities that are necessary for this proposed measure are strongly influenced by the actual network usage, the topology of the network and the storage time frame. Nevertheless, this is mitigated by the fact that the proposed measure only requires the storage of simple textual information, which is usually only a few hundred bytes, as opposed to complex binary information, such as images. Furthermore, the amount of storage required is reduced to the defined time frame of data storage and can be further cut down by stripping the network transaction information down to only necessary information in the sense of its meaningfulness, as discussed earlier. Even if a complex network usually has multiple sub-networks and the storage is therefore required for more than one gateway server, this is quite negligible given the current prices of storage devices and the speed of their increasing capacity development (Coughlin, 2020). The storage itself does not affect the functionality of the systems.

15.5.5 *Outlines for a system of plausible assurance of non-involvement*

The previous conceptual and technical requirements analysis, as well as the analysis of already existing technical capabilities of IT systems and their network components, show that a system for the plausible assurance of non-involvement in a cyber incident could be established based on already existing IT networking components. As these provide all the necessary information and nearly all necessary tools, the following aspects need to be taken into account during implementation:

- Ensure that information about network connections is collected and gathered on all relevant network and subnetwork gateway servers.
- The existing logging system is – if not already in place – extended by storage of these logs for a sufficient amount of time.
- The storage is performed tamper-proof, for instance, via immutable data storage technology and keeps the balance between the required secrecy of the actor collecting information as well as necessary details for a potential accusing actor.
- The storage remains under the control of the establishing actor and is managed by the IT service personnel that already runs the IT security measures of the affected IT networks.

Regarding the cyber incident model presented earlier as well as the model of common cyberattacks (as outlined in Figure 15.1), the de-escalation measure could work as follows:

- After the entitled authorities of Actor (B) detected and analyzed the incident coming from an IT system within the networks of actor (D), they request information about the gateways of this specific network N(D).
- If actor (D) is cooperating and has the logging mechanism in place, it detects unlawful access coming from within the networks of actor (A), which can be linked to the cyberattack activities against actor (B). It provides this information to actor (B).
- As actor (B) suspects actor (A) to be the origin of the attack, it requests information from actor (A) about all connections for the time frame of the attacks on N(B) for all gateway servers of N(A). This suspicion is usually based on overall political circumstances, bilateral political tension, recent or former events and also influenced by political considerations of the accusing state (B) (Broeders et al., 2020). If there are multiple suspected actors, these steps would need to be taken with them accordingly.
- Based on the information provided by actor (A)[2] and potentially supplemented by additional information from actor (D), actor (B) checks for connections from N(A) to N(D) that match the attacks in time and scale as seen from N(D) to N(B) to trace back the full attack path. This matching can be performed highly effectively by simple text search and comparison algorithms in no time to at least look for indicators of the presumption of actor (B). There also exist tools capable of visualizing network connections[3] to help human analysts quickly get an impression of the situation. As actual cyberattacks require many network

[2]It is important to mention that the requested information by actor (A) has to get transferred to actor (B). The time required for this is, as already mentioned, very dependent on the networks of actor (A) as well as on the transmission capacities and is an important factor despite the very effective data storage. This can be estimated in the case of a concrete implementation and on the basis of network information then available and taken into account in the political processes. Alternatively, however, direct access by actor (B) to actor (A)'s databases would also be technically possible. Even if this is associated with increased political costs, it could offer a way out in the extreme case of a danger that needs to be averted immediately.

[3]See e.g. this list of open-source tools for network monitoring and traffic visualizations, especially the tool "Nagios Core". https://www.comparitech.com/net-admin/open-source-network-monitoring-tools/

connections, an actual attack by actor (A) would be immediately visible in the data set. If no indicator is found to support this assumption, the information must be checked again more closely to rule out the possibility that connections have been overlooked.

- A de-escalation of the exemplarily assumed tension would be achieved if the data provided by actor (A) either contain no information about outgoing connections from its networks N(A) directly towards the attacked system of actor (B) for the time of the attack and if the connections from N(A) towards the neutral networks N(D) cannot be linked to the attacks against N(B).

The data collection of actor (A) is highly dependent on its IT network structure and the institutional organization that is in charge of operating the gateway server. However, as mentioned above, the information is usually already logged and is available in structured log files (usually one file per day per gateway) due to common IT security measures. This means they can be collected and presented immediately if the proposed measure is in place and the organizational structures are working. The data sharing between the actors can either be realized via secret communication channels between (A) and (B); be published if this supports the political signaling effect of the accused actor (A); or (B) could create public pressure that requires this step by publicly requesting the information from (A). As network log data is only text-based information, log files are rarely larger than a few gigabytes (1 gigabyte can store 1 billion characters) and transmitted without any delay within minutes. The provided data could either be validated and checked by actor (B) as demonstrated above, or by a neutral party whose conclusions would be accepted by both actors, such as a UN institution. Beyond this conflict scenario, an attacked state could also decide not to go public with the incident, not blame another state, or request information. In these cases, our proposed measure would not provide any support. This said some experts argue that transparency in political relations is not always the best solution. However, following this line of argument would go beyond the scope of this paper (J. M. Brown & Fazal, 2021).

The above-presented outline is not to be understood as a ready-made blueprint for a measure to be implemented immediately. Rather, it is a theoretical concept. However, the concrete considerations depend to a large extent on the specific circumstances in the individual networks (e.g., which IT systems are used), including questions about technological developments (e.g., the storage space required and, therefore the required additional IT hardware to store the logged information changes quickly, so that a certain fixed reference becomes obsolete very quickly (see e.g. Coughlin, 2020). Such considerations also depend on the political considerations of each implementing actor (e.g., for what time period logs are stored). Additionally, a real-world attack could involve many different proxies and an attacked state could suspect multiple states to be the origin. Nevertheless, the immediate attribution of attacks is largely a matter of political considerations and is based on previous events and situations of tension. However, even if further analysis of these parameters and their interconnection is valuable, this goes beyond the scope of this work.

15.6 Discussion and Outlook

15.6.1 *Limitations and potential pitfalls*

The developed procedure faces some potential pitfalls that will be discussed in this section. The most relevant limitation is the measure's limited use case for interacting states and their necessary motivation to join and comply with the measure. Section 15.5.1 has already discussed this prerequisite, pointing out that a state's participation and compliance to the measure directly influence the state's credibility and thus, its possibilities to reduce the risks of the escalation of cyberconflicts. Regarding implementation, there are various ways to cheat, as will be discussed below. Nevertheless, the main motivation for states to prevail from cheating is the plausibility of the information provided and the effectiveness of the measure in the long term.

Use of proxies

In terms of the plausibility of the measure and the information it can provide, a particularly critical limitation is based on the use of proxies for actions of military or intelligence services in cyberspace. A proxy could either be IT systems within the state's jurisdiction that are not associated with the state-owned military or intelligence services and are not officially under its control. Russia, e.g., is known to use so-called patriotic hackers, which are officially civilian groups that are suspected of being under the unofficial influence and command of Russian national intelligence services. However, a proxy could also be one or more IT systems used to perform cyberattacks that are located in another state. This scenario is quite common and one of the previously discussed reasons for the risk of miscalculation. Unfortunately, there is no technical solution against states acting contrary to agreements and avoiding attribution in this sense (Wingfield & Wingo, 2021). To a certain degree, a strong civil society could reveal such behavior, and international intelligence services cooperation of states could help to uncover such operations by sharing information or – in the best case – the proxy system is controlled from IT systems where the proposed measures is in place and would leave traces in the logged information. Besides this, cyberattacks carried out by non-state actors without the direct or indirect involvement of state institutions cannot be mitigated by the proposed measure. This is an inherent limitation, as only IT networks directly under the control and management of military forces or intelligence services are to be logged. Any logging beyond this would pave the way for censorship or surveillance, which is neither desired nor intended by this paper. In the best-case scenario, a state commits to the due-diligence principle of state liability and prevents malicious activities that are performed under its jurisdiction. However, this goes beyond the scope of this work. Another aspect might be that a state uses the commitment to the proposed measures as a pretense to establish complete civilian surveillance. Apart from the fact that such a state would probably use any circumstances to justify such an undertaking (like fighting criminal or terrorist activities), this is neither desirable nor necessary or useful for the proposed application context.

Dishonest participation

A participating actor could decide to avoid attribution by not logging all relevant information and, e.g., by leaving out incriminating connections. Although defining rules that omit some connections is highly system-dependent, this can easily be achieved either by excluding a specific range of IP addresses to not get logged or even specific for specific types of transferred data. In addition, an implementing state could also decide to leave out some gateways completely, which we have previously referred to as hidden channels. This could be mitigated to a certain degree by immutably logging the logging rules themselves, which can then be verified against transmitted information to check for inconsistencies by using special verified logging systems, such as the so-called Trusted Execution Environment (TEE) (Felton, 2019). However, this kind of cheating cannot be completely prevented because as long as the logged information is under the control of a state, it is possible for it to cheat. On the other hand, this method of non-compliance can be detected when an attacked state sees network connections from the accused state in its own log files that did not show up in the logged information presented by the accused actor. This discrepancy between the provided information, on the one hand, and the information of the attacked state, on the other hand, can be revealed immediately with a simple 1:1 check of both data sets. If an attacker uses proxies to perform the attack or if the attacker uses obfuscation measures to hide the attack path while using the proposed measures to signal non-involvement, this form of cheating is also not fail-safe. Though complex and time-consuming, the process of attribution and analysis of attacks in the aftermath can probably reveal the origin of an attack or at least hint to the performing state, which – if it contradicts the initially provided proof of non-involvement – will undermine the state's credibility. In any case, although this dishonest behavior may benefit a state in the short term, as it seems to prove its alleged non-involvement in the incident, in the long term, it undermines plausibility and credibility and renders the measure and its de-escalating effect useless. Each implementing state must therefore decide which path to follow. Finally, it must be recognized that states and their institutions often have divergent, sometimes contradictory interests and that political decisions and intentions sometimes contradict concluded agreements. From a pragmatic perspective, although such behavior undermines the value of a specific attempt to de-escalate, it does not undermine the value of the proposed approach in general, given the tense situation in cyberspace, the high potential for misinterpretation and the necessity for peace-sustaining measures.

Limited feasibility and range of applicable networks

As discussed, the radius of a possible implementation is limited to specific networks like military and intelligence services networks to prevent the establishment of a surveillance system. This is not considered problematic in terms of the research question, as the measures directly aim at the implementation by government institutions, which already have a special role that is usually associated with high responsibility and legal obligations. Although highly dependent on the political will, establishing the proposed measure within their networks is therefore considered applicable and legally indisputable if the political will to provide de-escalation measures is given. Furthermore, in most cases, military and intelligence operations follow specific orders, strategic and tactical planning, and have a chain of command and management of their activities. In particular, cyber

operations are often the result of years of planning, building technological resources and personal know-how, and therefore do not occur isolated and out of the blue. Such long-running activities often use a continuously used channel to collect sensitive information, and the backdoor to the attack system is kept active, well hidden and up to date with the potentially evolving security measures in the target system. This highly increases the probability that at least some traces of network activities are logged as indications of harmful activities in the named systems for the time frame of the storage measure (as discussed in section 15.5.4), even if the attack is carried out indirectly or via proxies.

Limitations of the logged information, secrecy and confidentiality

A double-edged aspect is an extent of collected, stored, and potentially committed information about network activities. On the one hand, more detailed logging has a higher informative value regarding the intended effects, but on the other hand, it might contain secret information that can prevent actors from establishing such measures. This can be diminished by anonymizing the stored information from a level that would allow identifying a unique IT system to a level where only the fact that the connections originate from the network would be logged. In addition, the information remains secret with the actor that deployed the measure until needed and is deliberately used only to prevent an imminent crisis with political tensions necessitating such means of de-escalation. Besides the already discussed relevance of a sufficient time frame of data storage and the importance of including all relevant gateway servers that directly influence the credible argumentation, another limitation is given when cyber activities involve anonymization services, such as the TOR (an abbreviation for The Onion Routing) network (The Tor Project, 2019). Such services remove any information from packages that allow their assignment of an endpoint of a connection to its origin. Even if this effectively undermines the approach of attributing cyberattacks to perpetrators, the proposed measure can nonetheless provide plausible information that there were no connections between the networks of the accused actor and the servers of the anonymizing services during the specific time frame of the attacks.

Another limitation concerns the aspect that the data, which can show actor (A) was not involved in a specific attack against actor (B), could at the same time contain critical information that would reveal cyber activities committed by (A) against another actor (C) that, until then, would not have been discovered. This could potentially discourage actors from implementing this measure but can partly be mitigated by the mentioned anonymization measures. In addition, the signatures (e.g., the amount and timing of network connections, as well as the extent and type of transferred data) differ between cyberattacks. This means that even if the provided data revealed ongoing cyber operations that are not part of the actual incident, the logged information and its characteristics could provide plausible information to argue in favor of the accused state.

Necessary technical and organizational adjustments

At last, it needs to be pointed out that the proposed measures need an adjustment of existing IT network infrastructures with an extension of the necessary processing and storage capabilities as well as the associated costs to sustain these capabilities. However,

as argued above, such capabilities often already exist for IT security measures, thus limiting the need for complex IT infrastructural changes. Nevertheless, any additional hardware and software need maintenance and skilled personnel. As the storage could contain sensitive information, implementing the measure could also require the establishment of technical access control mechanisms alongside the organizational structures and permissions. With regard to the required storage capacities, it has already been argued that it highly depends on the actual IT network topology, size and activity. Nevertheless, as the proposed measure is thought first and foremost as a political advance and needs to be backed up by national legislation, it is worth pointing out that the actual technical requirements have not played a role in the past in similar legal initiatives such as data retention (EU-FRA, 2017) and were considered a necessity for IT providers to adapt.[4]. Section 15.6.3 will, nevertheless, discuss how simulations could help to estimate the technical dimension.

15.6.2 *Conclusion*

The technical analysis presented above introduces a system of network connection logging and tamper-proof storage that enables an actor to provide network activity information that can plausibly assure their non-involvement in a given cyber incident to an accusing actor or a neutral intermediary. The analysis shows that, besides the political credibility of an accused actor, the keys for any line of argument depend on completely establishing the logging mechanism on all relevant gateway servers, on the time frame in which logging data is stored and kept, and on its tamper-proof acquisition and storage. The possibility to anonymize logged information to the network level without losing its evidential value provides the technical requirement of secrecy as well as making the measure ready for upcoming IT security measures like MTD. The conceptual outline shows that such systems can be built upon existing technologies and often already available IT hardware while including a few adjustments relating to storage capacities and maintaining the capacities to keep up with newer developments like the current shift of the IP. To create incentives for implementation, the measure of providing data allows different approaches that can disguise irrelevant and sensitive information. Despite the presented limitations, the proposed measure can provide a significant tool to circumvent the inherent problems of cyberattacks of missing data, their interpretation, miscalculations, and their attribution. With regard to the overall goal of arms control to provide tools for reducing the risk of armed conflicts, the measure can support this objective and provides a tool that allows overcoming the current insecure and unstable status quo.

In the examples of cyber incidents presented, it became clear that a state's perception of an imminent cyber threat and the need to respond to it can lead to escalation. The analysis of the attribution problems highlighted how the ambiguity of data could lead to miscalculations regarding the scope, the origin and the intention of a cyberattack –

[4]A rough indication of the costs required by this measure is provided by an estimate of the German "Bundesnetzagentur" (Biermann, 2015; Bundesnetzagentur, 2015). In 2015, in the course of the discussion on the introduction of data retention, in which connection data should be retained for 10 weeks and location data for 4 weeks, they estimated the costs for telecommunications companies at approx. 100,000 euros for companies with up to 1000 telecommunications subscribers and up to 400,000 euros for companies with up to 30 million telecommunications subscribers. The estimates were based on surveys of German ICT companies.

be it espionage that has gone wrong, sabotage, or an actual open, disruptive attack – and therefore lead to likely misresponses. This analysis also shows that the potential for escalations exists even when there has been no actual cyberattack but merely a fear of an imminent attack on critical systems or a perception of preparation for an attack, when there are no means for involved actors to provide information that can preclude this, e.g. by demonstrating that no cyber activities that usually precede such attacks have been taken. This fear has been fuelled by ongoing successful or attempted cyberattacks against critical infrastructures over the last years (Weinberg, 2021) and the recent demonstration of some states not prevailing from such prohibited measures. Such political situations with smoldering conflicts, a high degree of mistrust, and political and military tensions exist, for instance, between Pakistan and India (Baezner, 2018; Hess, 2021). These tensions have been affecting regional civil communities in the border regions for decades, while political communication channels are scarce. For such situations, the proposed measure can potentially provide information that can be used to de-escalate a tense political situation by showing the absence of a cyber threat.

Concerning the context of this paper and the research question, the proposed measure can provide means to reduce the risk of conflict escalation in the aftermath or during ongoing cyber incidents due to misattribution or misinterpretation of information, thus helping to refrain from the use of force (UNIDIR, 2018). The measure is also capable of providing the required degree of transparency of current cyber activities within an actor's military networks to show that there is no imminent threat of malicious cyber activities. The necessary degree of commitment of involved actors recommends such measures for situations between actors with a high degree of political tension, where no other communication channels for crisis reduction exist, which relates to the commitments for stronger international cooperation in cyberspace (UNGGE, 2015b). With regard to the consequences for personal rights and data privacy, the proposed measure should be limited to an application in highly critical scenarios only and only for potentially affected networks of military forces or intelligence services. Notably, the stakes for private, commercial and public IT systems, their protection and integrity and, respectively, for the communities that rely on them are high; and de-escalation measures are therefore strongly needed. Even if only established unilaterally, a state's transparency, credibility, and self-restriction of its own capabilities to conduct offensive measures in cyberspace could, as the analysis shows, be a valuable signal to overcome distrust (CSCE, 1975) between potential conflict parties. This does not reduce the offensive cyberspace capabilities of other states in question, but would – with the restriction of the discussed limitations – prevent the hidden, undetected use of a cyberattack in foreign IT systems as every connection would be logged. The proposed approach should help to foster the important task of preventing conflict escalations.

15.6.3 *Outlook*

Given the aspect that this paper presents a concept and outline but not an actual implementation proposal, a next step could provide such a proof-of-concept system. An actual implementation would have to be based on an analysis of the current military and intelligence service IT networks, their structures and technical characteristics. Such an analysis could therefore be used to evaluate the exact dimension of the technical parameters mentioned, such as storage duration, necessary additional storage space, etc.

This would provide a basis to investigate the dependencies of the technical parameters, the necessary level of detail of the provided information as well as their impact on the intended plausibility presentation. In addition, future work should also perform simulations and calculations of how different network sizes and topologies, network traffic capacities and storage duration influence the necessary storage capacities and, thus, the necessary new infrastructures. Following a proposal for international accountability in cyberspace (Davis II et al., 2017), the collected information of this measure could also directly help to strengthen the international credibility of the attribution processes carried out by a supranational institution under the UN regime (Davis II et al., 2017; Droz & Stauffacher, 2018). Tamper-proof information that is collected in a standardized procedure could provide relevant contributions to this task. With a long-term perspective of arms control and arms regulation for cyberspace, the approach might also be implemented in safeguard agreements. Such treaties of international security could obligate member states to collect and share credible information among themselves that can be used for mutual compliance control if certain restricted cyberweapons have not been used against each other or used at all. Besides the proposed implementation as a de-escalation measure, it can also be evaluated how the measure can be implemented between allied states in order to jointly prevent false-flag operations among this group of states and their IT systems and networks. Further research could analyze whether and how traces of malware samples or logging information collected during cyberattacks from third parties could be used and compared against logging information that has been collected by the proposed measure and provided by a state to detect compliance when implemented as a safeguard measure. Such a comparison could offer additional tools to verify that an attack has been allegedly performed by an actor over the detected, accused third party and to further reduce the possibilities for hidden attacks. Additionally, the proposed approach could be extended to actors whose IT systems have been verifiably used for cyberattacks to plausibly argue that these had been performed by external hackers who misused the IT systems. This could provide a relevant forensic approach to bypass the current third-party-based hacking methods that are commonly used. Such third-party attacks often use public IT systems, this includes the challenge of how these IT systems could be integrated into the proposed de-escalation measures while preserving the integral principle of data protection, privacy and civil rights. Another issue could address the minimization of the proposed data storage, either in terms of reducing necessary resources or – more importantly – in terms of secrecy. This can be performed, for instance, by differentiating the storage of data connections into separate lists of addressed networks and connection metadata like application types. These lists could enable an accused party to provide precise data for specific incidents and prevent the handover of excessive, irrelevant, or potentially secret information. Finally, the measures could also be used to strengthen the development of digital trust and confidence-building measures as well as verification regimes that monitor and control the compliance of actors, e.g., towards a hypothetical non-use agreement of cyberweapons or an agreement to not attack specific critical infrastructures. In this context, it would be necessary to develop and establish practical control measures such as on-site or live inspections of gateway servers by neutral third parties in the sense of the safeguard agreements performed by the International Atomic Energy Agency (IAEA) to control the nuclear program of Iran or the verification regimes performed by the Organization for the Prohibition of Chemical Weapons (OPCW) under the Chemical Weapons Convention (CWC) – with the difference that inspections would affect military or intelligence service facilities in this case. Even the unilateral implementation of the proposed system, as a means of self-restraint from conducting offensive covert military

or intelligence operations and, notably, as a means to signal transparency regarding one's own cyberspace activities, could establish a strong political signal of trustworthiness and political willingness. As cyberspace is increasingly becoming a domain of military power play, such signals are urgently needed.

16

EXTRUST: REDUCING EXPLOIT STOCKPILES WITH A PRIVACY-PRESERVING DEPLETION SYSTEM FOR INTER-STATE RELATIONSHIP

Abstract Cyberspace is a fragile construct threatened by malicious cyber operations of different actors, with vulnerabilities in IT hardware and software forming the basis for such activities, thus also posing a threat to global IT security. Advancements in the field of artificial intelligence accelerate this development, either with artificial intelligence enabled cyberweapons, automated cyber defense measures, or artificial intelligence-based threat and vulnerability detection. Especially state actors, with their long-term strategic security interests, often stockpile such knowledge of vulnerabilities and exploits to enable their military or intelligence service cyberspace operations. While treaties and regulations to limit these developments and to enhance global IT security by disclosing vulnerabilities are currently being discussed on the international level, these efforts are hindered by state concerns about the disclosure of unique knowledge and about giving up tactical advantages. This leads to a situation where multiple states are likely to stockpile at least some identical exploits, with technical measures to enable a depletion process for these stockpiles that preserve state secrecy interests and consider the special constraints of interacting states as well as the requirements within such environments being non-existent. This paper proposes such a privacy-preserving approach that allows multiple state parties to privately compare their stock of vulnerabilities and exploits to check for items that occur in multiple stockpiles without revealing them so that their disclosure can be considered. We call our system ExTRUST and show that it is scalable and can withstand several attack scenarios. Beyond the intergovernmental setting, ExTRUST can also be used for other zero-trust use cases, such as bug-bounty programs.

Original Publication *Reinhold, T., Kühn, P., Günther, D., Schneider, T., & Reuter, C. (2023). EXTRUST: Reducing Exploit Stockpiles With a Privacy-Preserving Depletion System for Inter-State Relationship.* IEEE Transactions on Technology and Society. *https://doi.org/10.1109/TTS.2023.3280356*

16.1 Introduction

The threat of malicious cyber activities is omnipresent and state actors are becoming an increasingly important part of this development (Giles & Hartmann, 2019), either due to the progressing militarization of cyberspace (Koch & Golling, 2019; UNIDIR, 2013) or due to cyber espionage operations (Buchan, 2018; Georgieva, 2020). At the same time, advancements in the field of artificial intelligence (AI) are being used to automate cyber defense measures (Dhir et al., 2021), to develop AI enabled cyberweapons (Reinhold,

2021e), or to detect and predict software threats and vulnerabilities (Amarasinghe et al., 2019; R. L. Russell et al., 2018). In particular, knowledge of vulnerabilities is an integral part in most of these cyber operations in order to breach foreign IT-protection measures, and intelligence services and military forces stockpile such critical information without disclosing it for rectification (Ablon & Bogart, 2017; Rovner, 2020). However, any serious and capable exploit withheld by a state for its own purposes becomes a potential threat for everyone, including the state itself, its economy, and civil society (CCDCOE, 2020) as the exploit *EternalBlue* exemplified in 2017 (Cimpanu, 2019; Schulze & Reinhold, 2018).

One way out of this dilemma is a so-called vulnerability equity process (VEP) (Milch et al., 2020), an institutionalized measure to regularly assess the criticality of stockpiled exploits and vulnerabilities to (re)consider their disclosure that could take place under the leadership of extra-national entities, such as the United Nations (UN) (Schulze, 2019). A major obstacle for such an approach is the reluctance of participating parties to disclose sensitive information about their own capabilities, which is generally seen as giving up tactical advantages, effectively resulting in an international arms race for offensive cyber capabilities (Harknett & Smeets, 2020). Historically, such situations have been countered by efforts to reach mutual agreements between states on arms control and reduction measures, *i.e.*, treaty-based agreements to limit the risks of proliferation of weapon-enabled technology, to prevent its use with potentially disastrous consequences, or to reduce the risks of conflict arising by mistake or technological failures (Reinhold & Reuter, 2019a). Although political approaches for cyberspace have been proposed by the UN (Rõigas & Minárik, 2015; UNGGE, 2015a), the OSCE (Security & Europe, 2016), and other organizations and important steps towards an effective cyber arms control, like the exchange of threat information (Kuehn et al., 2020; Sauerwein et al., 2017), have been established, this is not suitable for limiting or reverting the aforementioned international cyber arms race of vulnerability stockpiling. So far, no proposal focuses on this specific challenge and the particular constraints of state actors, with their requirements of confidentiality, their potential mutual mistrust, and individual security concerns (Reinhold & Reuter, 2019b).

Our Contributions. In this paper, we propose a technical solution called ExTRUST based on a multi-party computation approach that allows multiple actors to compare vulnerability stockpiles for matching entries while preserving their confidentiality. This includes an approach for the unique machine-readable identification of exploits that allows them to be checked for matches. Our solution is designed for a zero-trust environment and does not rely on any preconditions of trust in advance or assumptions of good nature. This contributes to the development of measures for an international agreement to deplete vulnerabilities while circumventing the problems and impediments of intergovernmental cooperation.

Beside this contribution, this paper further aims to provide an example of how politics is – sometimes – in need of technical solutions, in this case even for challenges regarding international security. As computer scientists and engineers are the experts on the domain of cyberspace, shaping it by developing software or even defining its constraints and rules themselves, we would like to encourage taking the responsibility that this entails seriously and support the peaceful development of this globally shared domain.

Section 16.3 analyzes the requirements of ExTRUST both on a conceptual and IT security level. Section 16.4 discusses how vulnerabilities can be uniquely described in a machine-readable form and presents our approach that allows to check for matches. Section 16.5 presents an introductory section that exemplifies the intended system and the requirements for using a Blockchain-based prototype approach for ExTRUST and discusses challenges. Section 16.6 presents our contribution of a privacy-preserving exploit depletion system for zero-trust relationships using multi-party computation. Section 16.7 discusses the approach and evaluates it against the requirements. It also presents different application scenarios beyond state actors. Section 16.8 concludes this paper and provides directions for future work. In order to maintain readability, the technical details can be found in the Annex in Section 16.9.

16.2 Related Work

Since our paper covers and combines different computer science topics, this section summarizes the existing work on malware identification (section 16.2.1), vulnerability mitigation (section 16.2.2), and promising cryptographic protocols (section 16.2.3). Based on these descriptions, the research gap is described (section 16.2.4), which is closed by our approach.

16.2.1 *Vulnerability Terminology and Malware Identification Methods*

An important prerequisite for comparing exploits – as the core of a depletion system – is the ability to create deterministic vulnerability descriptions. Early attempts were based on the creation of so-called malware *signatures* (Cohen, 1987), which function like a *fingerprint*. Current malware detection approaches use a different approach that is either based on the entire binary code of the malware, *i.e.*, the exploit and payload (Kirat & Vigna, 2015), or the compromised storage to create signatures (Petrik et al., 2018). Beside the actual detection of malware, other research area focuses on the description and identification of exploited vulnerabilities. The popular national vulnerability database (NVD) provides a semi-structured database of known vulnerabilities (MITRE Cooperation, 2023), however, Dong *et al.* (Dong et al., 2019) showed that the NVD entries are inconsistent compared to other vulnerability databases. Compared to the common vulnerabilities and exposures (CVE), the NVD entries differ in their announced project names or versions. Alternative approaches were introduced by Sadique *et al.* (Sadique et al., 2018) with the *Structured Threat Information eXpression (STIX)* and the *Vocabulary for Event Recording and Incident Sharing Framework* (VERIS) (VERIS, 2019) that can be used to describe, share, and publish threat information. Both definitions, STIX and VERIS, offer a syntax for different types of threats, including malware, exploits, and vulnerabilities. Some entry fields in NVD, STIX, and VERIS may contain unstructured information that undermines unique descriptions. Martin *et al.* (Martin et al., 2011) propose the common weakness enumeration (CWE), a dictionary of weakness classes that can be used to classify vulnerabilities, an approach we use in section 16.4 to identify vulnerabilities.

16.2.2 *Vulnerability Mitigation & External Depletion Measures*

Vulnerability research and mitigation methods have been a topic in IT security for several decades (Carlini & Wagner, 2014; One, 1996; Shacham, 2007; You et al., 2017). One measure are so-called *bug-bounty programs* (Zhao et al., 2014) like e.g. *HackerOne* (Perlroth, 2015), which aim to attract IT security practitioners to penetrate advertised systems and services and report loopholes in software or services. Other programs are run by Mozilla, Facebook, and Microsoft (Facebook, 2018; Mozilla, 2017; Zimmerman, 2017) or *Project Zero* (C. Evans, 2014) by Google, which focuses on the search for zero-day vulnerabilities. These programs, which we further refer to as *external depletion measures,* aim to identify vulnerabilities in popular IT products in order to disclose them to the producers and get them fixed as a depletion measure.

In contrast, *internal depletion measures* that focus on an actors secret exploit stockpile of already known, but not yet disclosed vulnerability information have not yet been proposed, especially not for the given application context of interstate cooperation and international security. Practical approaches at this international, intergovernmental level have so far been limited to transparency and confidence-building, rather than arms control and the non-proliferation or disarmament of malicious cyber tools. (Reinhold & Reuter, 2019a).

16.2.3 *Cryptographic Protocols*

Our ExTRUST system is related to well-studied cryptographic protocols like multi-party computation (cf. section 16.2.3), private set intersection (cf. section 16.2.3), and trusted hardware (cf. section 16.2.3). These approaches are further elaborated in the following.

Multi-Party Computation (MPC)

The first approaches to multi-party computation (MPC) of functions represented as a Boolean circuit were proposed by Yao (Yao, 1986) for N 2 parties with constant round complexity, and by Goldreich, Micali, and Wigderson (GMW) (Goldreich et al., 1987) for any number of parties N with round complexity linear in the depth of the Boolean circuit. Beaver, Micali, and Rogaway (BMR) (Beaver et al., 1990) extended Yao's protocol to the multi-party case while maintaining the linear round complexity. Based on this initial work, many research projects followed, showing the practical feasibility of MPC for many privacy-preserving applications, such as auctions (Bogetoft et al., 2009), set intersection (Pinkas et al., 2018), and machine learning (Mirhoseini et al., 2016). Kamara *et al.* presented an outsourcing technique (Kamara et al., 2011), which allows N parties to outsource the MPC protocol to $n \ll N$ parties.

Private Set Intersection (PSI)

Private Set Intersection (PSI) has been proposed to identify malware (cf. section 16.2.1) in a single client and server environment (Kiss et al., 2017). A recent survey and performance comparison of different PSI protocols by Pinkas *et al.* (Pinkas et al., 2018) demonstrates that the approach proposed by Pinkas, Rosulek, Trieu and Yanai (Pinkas et al., 2020) is currently the fastest PSI protocol which can handle malicious security. In our proposed application context, we have multiple parties, hence we are mainly interested in multi-party PSI. Multi-party PSI protocols with passive security are applied by Kolesnikov *et al.* (Kolesnikov et al., 2017) and Inbar *et al.* (Inbar et al., 2018). A scalable, maliciously-secure multi-party PSI protocol is presented by Hazay and Venkitasubramaniam (Hazay & Venkitasubramaniam, 2017). Huang *et al.* (Huang et al., 2012) use a general MPC framework to privately compute the set intersection between two parties.

Trusted Execution Environment (TEE)

Another promising approach for a privacy-preserving exploit depletion system is to securely isolate the execution into a *trusted execution environment* (TEE) (Anati et al., 2013), that allows untrusted data to be computed in a secure environment that is isolated from all other executions running on the same machine, where it is protected against manipulation and disclosure. TEEs are omnipresent in all Intel processors from the 6th generation upwards as *Intel Software Guard Extension* (SGX). Although many works use Intel SGX for efficient secure multi-party computation (Bahmani et al., 2017; Felsen et al., 2019; D. Gupta et al., 2016; Koeberl et al., 2015; Küçük et al., 2016), TEEs are not suitable for applications when states are involved, since this would require that state actors trust the hardware-producing countries not to manipulate the TEEs, *e.g.*, by including backdoors.

16.2.4 *Research Gap*

Above all, practical measures are a mandatory aspect of potential arms control and disarmament treaties, as history and insights into former weaponized technologies have shown (Goldblat, 2002). Existing IT methods such as Multi-PSI (Hazay & Venkitasubramaniam, 2017) (cf. section 16.2.3) and secure hardware (Felsen et al., 2019) (cf. section 16.2.3) have not been applied to exploit depletion, especially regarding the demands and particular constraints of an interstate zero-trust environment. Such a protocol for pairwise PSI among N parties, as required for a privacy-preserving exploit depletion system, is currently not available. Thus, our approach ExTRUST proposes a Boolean circuit that implements the desired functionality via MPC (cf. section 16.6).

16.3 Requirements Analysis

In this section, the requirements of ExTRUST are analyzed as a system for reducing exploit stockpiles, resulting from the chosen context of interstate relations. This list is divided into conceptual requirements derived from the specific constraints of the context of arms control, as well as the IT security requirements in combination with the selection of the adversary model.

16.3.1 *Conceptual Requirements*

As mentioned above, this paper focuses on cases in which two or more parties stockpile vulnerabilities and exploits. This reflects the character of arms control treaties, whose "practical" part of active mutual control or (limited) cooperation measures are always based on bi- or multilateral agreements (Reinhold & Reuter, 2019c) between a small group of states. Based on a rational choice consideration (Zangl & Zürn, 1994), our approach builds upon the following premises that we consider to be reflected by states that stockpile vulnerabilities (Kraus et al., 2019), as they resemble the considerations behind a vulnerability equity process (Milch et al., 2020):

- A state is aware that withholding a vulnerability poses a potential threat to its own IT systems.
- A vulnerability that is also known to other states is more likely to be considered a candidate for disclosure by a state, because
  - its intended effect is probably ineffective, but at least uncertain, as every other state that is aware of this vulnerability has probably secured its IT systems accordingly;
  - disclosing the vulnerability results in publicly available security patches that support the state's own IT security and also renders the vulnerability worthless for everyone.

On the other hand, all vulnerabilities are high-value assets for the stockpiling party. Given the context of state interaction, each party will try to avoid revealing any information that can lead to the loss of tactical advantages, while trying to extend these advantages by gaining information about the other parties. In addition, arms control measures are established in times of political tensions in order to avoid the outbreak of armed conflict. Based on these assumptions, we consider that ExTRUST has to operate in a zero-trust environment in which parties have to be incentivized to cooperate, while at the same time assuming that other parties are either extremely reluctant to disclose information, attempt to gain information for their own interest, or are even willing to cheat.

With these considerations in mind, ExTRUST aims to require as little cooperation as possible due to this zero-trust environment. This means that each party discloses only the absolutely necessary amount of information, thereby retaining all specific information about capacities and capabilities. Additionally, each party should be able to perform its

own check for intersections at any time without relying on further cooperation, dedicated data exchange, or any form of super-ordinate institution. Furthermore, information already provided should not be allowed to be altered, deleted, or corrupted.

In light of this context, the necessary measure needs to fulfil the following conceptual requirements (RC):

RC1 The measure has to enable parties to add information about vulnerabilities and exploits.

RC2 Intersection checks have to be able to be performed by either party at any time without having to obtain the consent of the other parties involved. A match is considered as such if at least two different participating parties have submitted identical information about vulnerabilities or exploits.

RC3 The system has to send feedback when it detects an intersection match.

RC4 Although real-time computability is not strictly necessary for processes that are usually politically slow, such as arms control measures, the system needs to be scalable with respect to the number of parties so that parties can join or leave at any time. While previous arms control treaties are usually established in a small circle of state actors that participate in mutual control measures, indicating there could be up to $N \quad 5$ participating parties in a real-world arms control scenario, this should not be the upper bound of our system.

RC5 The system should be operated decentralized and not require a specific neutral authority to operate or maintain the system.

16.3.2 *Adversary Model*

The two most common adversary models are semi-honest (passive) and malicious (active) adversaries (D. Evans et al., 2018). While semi-honest adversaries follow the underlying rules and procedures (in technical terms the so-called *protocol*) and try to extract as much information as possible from the transcript, malicious adversaries may arbitrarily deviate from the agreed rules. Given the zero-trust environment in the context of ExTRUST, we consider an active or malicious attacker as adversary model. Although technical security measures that protect against semi-honest adversaries are more efficient than those against malicious adversaries, we must consider state actors that might maliciously manipulate arms control computations and outcomes. Additionally, we assume a dishonest majority, *i.e.*, up to $N - 1$ parties may be malicious. The motivational scenario of ExTRUST is a highly security critical one in which top secret information may be exchanged. Hence, it should withstand several passive attacks, like eavesdropping, and also be shielded against active attacks, such as flooding or brute-force attacks. We have therefore chosen the model of the stronger adversary in contrast to the passive, semi-honest adversary. This decision also covers the application context of the zero-trust relationship between the actors involved.

16.3.3 *Technical and Security Requirements*

In addition to the conceptual requirements, the approach must meet additional security expectations to provide an applicable and secure measure of exploit depletion in a zero-trust environment. The requirements reflect the need for confidentiality and are important to motivate stakeholders to participate. These technical and security requirements (RS) are:

- RS1 The system must ensure the confidentiality of vulnerability or exploit information against any party.
- RS2 Submitted data should not be able to be withdrawn, modified, or corrupted by any party.
- RS3 The system needs to prevent false positive intersection results.

In the following, after discussing the identification of vulnerabilities as a necessary prerequisite of our system, we present a prototype solution for ExTRUST that addresses these requirements and illustrates its inherent challenges. Afterwards, we present our contribution of a MPC-based ExTRUST.

16.4 Identifier of Vulnerabilities

In this section, we propose a unique, machine-readable identification method for vulnerabilities to be able to match them. The mathematical description of the required properties and the associated challenges can be found in the Annex and are referenced here.

16.4.1 *Machine-Readable Vulnerability Identifier*

At its core, ExTRUST privately matches vulnerabilities or exploits of different parties. This requires using a vulnerability description method that results in the same machine-readable descriptions for the same vulnerability (section 16.9.1). An established approach to describe and thus identify vulnerabilities is provided by vulnerability databases like the NVD. The NVD's entries, for example, contain information used for identification. Their semi-structured format, however, makes it practically impossible for individuals to independently create the same identifier for a vulnerability. Therefore, we use the approach of Kuehn *et al.* (Kuehn et al., 2021) to achieve uniqueness, *i.e.*, we adjust the NVD's entry information by removing any free-form pairs and pairs that provide no information about the vulnerability itself (*e.g.*, the CVE-ID), align the structured information with the vulnerability descriptions, and add information about the vulnerable function, extracted from the vulnerability description.

The remaining fields are CWE and common platform enumeration (CPE) with the addition of the vulnerable function, which are structured and algorithmically comparable. The CWE (MITRE Cooperation, 2019) defines hierarchical layers of vulnerability

```
{
  "cpe": "cpe:2.3:o:tp-link:wdr7400_firmware:-:*:*:*:*:*:*:*",
  "cwe": 120,
  "fun": "copy_msg_element"
}
```

Listing 16.1: Vulnerability Identifier for CVE-2020-28877

weakness classes, while the CPE (Cheikes et al., 2011) provides a machine-readable way to describe platforms. If a vulnerability affects multiple platforms, we use separate vulnerabilities for each affected platform. The resulting vulnerability identifier is depicted in section 16.1 (for CVE-2020-28877).

16.4.2 *Analysis*

Using a simple object notation for the vulnerability identifier offers flexibility and extensibility, and by adding CPE and CWE as well as the vulnerable function as core elements, identifiers can be specific enough to create matching values when different actors describe and submit the same vulnerability or exploit. This is essential to identify matching vulnerabilities.

The main limitation of the vulnerability identifier's definition is based on a trade-off between the properties *accuracy* and *ambiguity*. Currently, it is still possible to describe two different vulnerabilities with the same identifier, or to use two different identifiers for the same vulnerability (section 16.9.2). This leads to false positives (two different vulnerabilities are mapped to one identifier) or false negatives (the same vulnerability is mapped to two different identifiers), respectively, depending on the level of detail implemented into the identifier. However, there are possibilities to adjust the identifier definition accordingly. Increasing the amount of information captured by the identifier makes the identifier more specific but introduces more ambiguity, *i.e.*, false negatives. Parameters to be added are the common vulnerability scoring system (CVSS) parameter information (*e.g.*, impact information) or the vulnerable path (*i.e.*, the filename in which the vulnerability resides) (Kuehn et al., 2021). Another way to adjust the identifier is the CWE's hierarchy depth. CWE classes are hierarchically ordered and thus offer generalization or specification. Including relations of the used CWE class increases the specificity of the identifier and could help to circumvent cases where identifiers use different CWE subclasses of the same top level class.

At this point, we want to stress that in the presented scenario (cyber arms control) false positives must be avoided, while false negatives are tolerable. If false positives are a common problem in such a system, it would drastically lose acceptance among states that are still interested in stockpiling vulnerabilities.

The size of the proposed identifier space is restricted by the number of CWE classes, the size of the CPE directory, and the possible function names, which serve as secret information. Individually, these spaces can be approximated in their size. For the space of possible function names FN, we assume a clean coding style, *i.e.*, function names are

descriptive and use at most three English words with any kind of connector (*e.g.*, camel case or underscores), which results in $\approx 2^{81}$ identifiers (section 16.9.3).

As argued, the presented approach is sufficient to describe vulnerabilities uniquely. It serves our needs with a trade-off in detail that avoids both different vulnerabilities being described by the same identifier, as well as the same vulnerability being described with different identifiers. Based on the current limitation of the identifier space, brute-force attacks remain a problem and efforts should be made to increase the identifier space. As an alternative to the proposed definition of identifiers, our system ExTRUST can work with any other scheme that is concise, structured, and unambiguous.

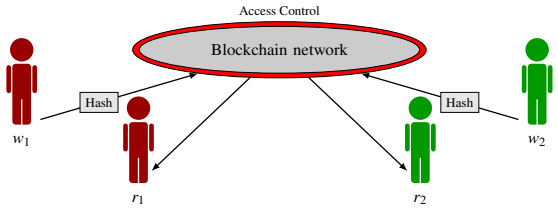

Figure 16.1: System architecture for BC-based ExTRUST. r_i and w_i denote readers and writers of actor i.

16.5 ExTRUST using Blockchain

In order to illustrate the challenges involved in implementing a privacy-preserving exploit depletion system, we have chosen a simple, straightforward prototype based on a Blockchain implementation, referred to hereafter as *BC-based* ExTRUST. Although this approach entails security flaws from a theoretical perspective, we want to use this prototype to illustrate, test, and analyze possible solutions regarding the requirements and the proposed depletion process, as an introduction for our multi-party computation-based approach presented in section 16.6. This section presents the architecture and proof-of-concept implementation of this prototype and concludes with a discussion of the requirements met as well as the identified constraints.

16.5.1 *System Architecture and Procedure*

In terms of conceptual requirements, BC-based ExTRUST should run in a *distributed setting with no central trusted authority*, with a *complete, secured, and tamper-resistant history* of all submitted information and should allow *asynchronous intersection checks* that can be performed by each participating party independently.

We have developed a prototype based on a private Blockchain technology (S. S. Gupta, 2017; Zheng et al., 2018) that provides all of these features. A private Blockchain is a distributed chain of *blocks* containing *transactions*, where each block references its previous block via hard-to-calculate mathematical challenges and cryptographic hashes to reference the block. This provides a tamper-proof history of all submissions, as any

modification would invalidate the adjacent entries. The data storage part of a Blockchain, the so-called ledger, is replicated to all participants and automatically synced between them. In private networks, access to it is walled by an access control manager (ACM). The interfaces for interaction with the ledger are called *smart contracts*. With regard to the system architecture, the ledger provides the storage space, the smart contract is responsible for the submission and comparison mechanism, and the ACM controls the access as well as the different layers of interaction permissions via roles and associated authorizations. To maintain the confidentiality of the submitted vulnerability identifiers, we secured the information using cryptographic hash functions (Katz & Lindell, 2014).

The overall procedure begins with the setup of the Blockchain instance (*nodes*) by each participating party and their interconnection to build an evenly distributed network. In order to submit a vulnerability, the vulnerability identification method we propose in section 16.4 is used to create an identifier for the specific vulnerability, which is then cryptographically secured using a hash function and finally stored in BC-based ExTRUST. Afterward, any participating party can perform a transaction, which checks for intersections between all hashes stored in the ledger and logs the output on the ledger. This way, a history of all actions performed is ensured, which is accessible to any involved party, including intersections. Nevertheless, parties that do not know the plaintext vulnerability identifier cannot obtain any information other than the fact that an intersection occurred.

16.5.2 *Implementation*

To focus on developing a proof-of-concept implementation of BC-based ExTRUST, we decided to utilize a private Blockchain framework (Davies, 2022) as it provides all relevant tools for the interaction of the actors with the system, the data structures for storing information, and all necessary data operations for reading, writing, and verifying information within the stored data. We have selected the *Hyperledger Fabric* (Voell et al., 2016) Blockchain framework because it is open source, actively maintained and well documented, and provides the rapid prototyping environment *Hyperledger Composer* (Hyperledger, 2017) with a boilerplate implementation for each part of the Blockchain network.

With regard to the permissions of participants using BC-based ExTRUST, we envision two roles: Readers, who can read the entire ledger and perform the transaction that checks for matching items; and writers, who can only submit items (see section 16.1). This restriction is only necessary due to the use of our chosen framework[1], otherwise, a party's submission may be intercepted and copied by other parties. The theoretical concept does not require this separation, because no party would be able to access other parties' information.

The items which are submitted and stored into the ledger are the vulnerability identifiers, as described in (section 16.4). As the plain vulnerability identifier must never be inserted into the Blockchain network to prevent its exposure to all parties involved, it is obscured before being submitted. We generate a cryptographic hash of a normalized JSON representation of the vulnerability identifier via `SHA3-512` (OpenSSL, 2020), following

[1] As framework, we chose Hyperledger Composer.

the NIST's policy on hash functions (NIST, 2015). This provides a 256-bit security level.

To interact with BC-based ExTRUST, the prototype system provides two transactions: The simple submission of hashed vulnerability identifiers and the transaction that checks the stored hashes within the ledger and triggers an event along with references to matching items, `checkIntersections` (section 16.9.4).

We want to stress that this prototype implementation does not yet take performance into account, as this is no core requirement of ExTRUSTand its proposed arms control application.

16.5.3 *Discussion of BC-based* ExTRUST

As indicated earlier, the development of IT measures is a novel approach in the field of technical tools for cyber arms control that has to balance conflicting objectives to a certain extent. For arms control, the aspect of minimum requirements for cooperation between the parties is essential, as it establishes the lowest possible barrier for participation. This is crucial for situations such as the intended one, in which trust cannot be assumed as a given motivation for cooperation. In previous treaties, this often meant a certain degree of pragmatism regarding the acceptance of "gray areas" and the possibility of non-compliance. The opposite objective is the requirement of technically secure solutions, as this too provides important incentives for participation. This in turn is likely to result in protocol specifications creating operational conditions that potential participants are not prepared to accept.

Considering the requirements, the Blockchain approach provides a manipulation-proof and distributed storage of all submitted information. Calculations are distributed and performed independently, thereby mitigating the need for a trusted third party to maintain the shared information, as well as any other form of cooperation beyond the actual submission. The system can include additional parties without adjustments or significant impact on the performance of the system, beyond the network capacity necessary to synchronize the stored information (Nadeem, 2019). In addition, the processing of submissions is not time-critical, which is considered a bottleneck for massive, high-traffic Blockchain applications (Chohan, 2019; Croman et al., 2016). By securing vulnerability identifiers, the confidentiality of the information is – at least theoretically – maintained both in submission and in intersection detection.

On the other hand, the Blockchain-based prototype has serious IT security issues, both for active attackers (like non-participating state parties that try to break the system in order to gain advantages and reveal secret information) and fraudulent, semi-honest state participants that try to gain information which goes beyond the agreed exchange. Notably, the information contained in the distribution of the ledger is vulnerable to brute-force attacks by testing hashes, as foreign countries could generate possible vulnerability descriptions and test them against their local ledger. The PSI literature has demonstrated that private elements cannot be hidden by simple hashing (Demir et al., 2018; Kales et al., 2019). The probability of creating an existing hash is based on the size of the identifier space and influenced by the number of its properties and values. As the identifier space

of BC-based ExTRUST is very small (28 bit, cf. section 16.4.2) brute-force attacks are very efficient and can be successfully exploited. In addition, the brute-force attack is completely local since states have a local copy of the whole ledger. Consequently, states would not even notice if a brute-force attack was exploited to find all ledger vulnerabilities. The brute-force attack can be slowed down (but not prevented) by using a difficult to parallelize hash function such as *Argon2* (Biryukov et al., 2016). Extending the identifier space for the vulnerability by more complex identifier descriptions is not an option either, as this increases the probability of describing the same vulnerability differently.

The Blockchain also faces other attack scenarios, such as the so-called 51% attack, which allows attackers to manipulate the ledger (Ye et al., 2018). Attackers could also use more subtle ways to create intersections to test foreign submissions by creating *fictional vulnerabilities* for rare software systems based on clever and informed guesses. This could also be used for targeted vulnerability suppression if a participating party creates and submits specific vulnerabilities, intentionally wrongfully signalling its possession in order to force the vulnerability to be disclosed. In addition, a dishonest state party could clone and resubmit hashes under its own flag, which would also cause BC-based ExTRUST to false signal to the original submitter that this particular vulnerability can be eliminated. However, such a cheat gives the attacker only a slight advantage, as they do not know what the cloned vulnerability information contains, and are likely to attract attention if performed regularly. A final IT security issue concerns passive adversaries that gain access to the ledger, as well as the complete disclosure of the ledger to non-involved third parties. Besides the brute-force attack, the attacker will be able to learn which hashes belong to which party via timing correlations, detecting the amount of different participating actors as well as the amount of submitted hashed items stored by each actor.

The BC-based ExTRUST prototype has shown that it provides the conceptual requirements that arise from the arms control context. Regarding the attack scenarios described, it is important to emphasize that for this application, any attempt to attack or misuse the system is contrary to the principles of the confidence-building aspect of such a mutual measure and its political signal of de-escalation. It is further expected that all parties comply with the defined rules to at least achieve a positive outcome for their own national security. Nevertheless, this expectation needs to rest upon a secure protocol that inherently prevents fraud and guarantees the promised confidentiality.

The following section presents the approach of MPC-based ExTRUST, an arms control measure that provides this level of security.

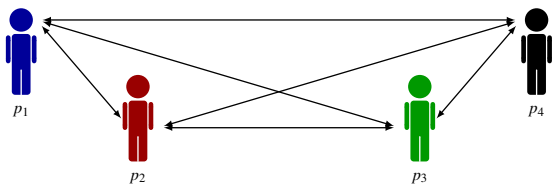

Figure 16.2: MPC setting with four participating parties $p_1, \ldots, p_4$.

16.6 ExTRUST using Multi-Party Computation

This section presents our approach for an MPC-based ExTRUSTto develop an exploit depletion system under the conditions of an untrusted environment that fulfills the discussed conceptual and security requirements (section 16.3), while avoiding the security problems that our prototype revealed (section 16.5). This approach is based on an interactive *Multi-Party Computation* (MPC) protocol as grounds for our MPC-based ExTRUST architecture. Secure MPC (Archer et al., 2018; D. Evans et al., 2018) enables N parties to securely compute a commonly agreed public function f on their respective secret inputs $x_1, \ldots, x_N$ without revealing anything other than the result of the calculated function $f x_1, \ldots, x_N$. MPC guarantees that each of the N parties will not learn any information (*e.g.*, input from the other parties or intermediate results of the computation) other than what a party would learn in the ideal world with a trusted third party. In the ideal world, all parties send their inputs $x_1, \ldots, x_N$ secretly to a trusted third party, which then locally computes the function $f x_1, \ldots, x_N$ and broadcasts the result to the N parties. In the proposed context of arms control, even if such a trusted third party existed (*e.g.*, in the UN framework), it would probably not be accepted by all state actors or, at the very least, would raise the barrier to participation in the proposed measure (see section 16.3.1).

In MPC, the function f that shall be computed is represented as a *Boolean circuit*. A Boolean circuit is a logical function whose operations are so-called *Boolean gates*. A Boolean gate takes a set of Boolean inputs (*i.e.*, either 0 or 1) and computes one Boolean output. We represent our MPC-based ExTRUST functionality as a Boolean circuit as efficient cryptographic protocols exist that can securely evaluate Boolean circuits. A Boolean circuit consists of inputs, outputs, and Boolean gates that have two inputs and one output in the Boolean set $\{0, 1\}$. The input of a gate can either be one of the inputs of the Boolean circuit or an output of a previous gate. In MPC, Boolean circuits usually only consist of `AND` and `XOR` gates, as any functionality can be realized using these two gate types. A two-input `AND` gate outputs '1' if both of its inputs are set to '1', while a two-input `XOR` gate outputs '1' if exactly one (but not both) of its inputs is set to '1'. For the actual algorithm that processes the submitted information and checks for collisions – the so-called *protocol* – we use the BMR protocol (Beaver et al., 1990) and refer to Braun et al. (Braun et al., 2022) for a detailed protocol description. We further use a well-established outsourcing technique (Kamara et al., 2011) to distribute the information processing for a group of N parties to $n \ll N$ parties. This setting for our MPC approach is shown in section 16.2. In summary, a subset of n from N (state) parties interactively run an MPC protocol on a Boolean circuit, which computes the functionality of ExTRUST. This setting allows us to evaluate the functionality of ExTRUST in a privacy-preserving manner, while reducing the number of active parties that are fully involved in the computation ensures the scalability of MPC-based ExTRUST.

The protocol requires that the parties have sorted their inputs locally before they are fed into the MPC protocol. In order to verify this, we use the Boolean circuit and open intermediate values so that the parties can abort the protocol execution if a malicious party has not sorted its inputs correctly. When opening these values, the remaining intermediate values before and after the opening process must be protected to allow further secure computation with them. Many efficient maliciously-secure MPC protocols

provide this property, known as *reactive MPC*, *e.g.*, (Bendlin et al., 2011; Damgård & Orlandi, 2010; Damgård et al., 2012, 2013; Nielsen et al., 2012). Apart from checking correctly sorted sets, we use reactive MPC to maintain the state of the secretly shared inputs after the end of a protocol run, so that submitted vulnerabilities do not need to be secretly shared again in the next iteration.

This MPC approach allows us to develop a privacy-preserving exploit depletion system that fulfills the requirements of the arms control context (section 16.3).

16.6.1 *System Architecture*

Our complete MPC-based ExTRUST architecture works as follows (see Fig.16.2): N parties try to find intersections of their own identified vulnerabilities between themselves and at least one other actor. These N actors securely evaluate a Boolean circuit (cf. section 16.2.3), consisting of `AND` and `XOR` gates (see Fig. 16.3), that takes as input the known vulnerabilities and exploits of the actors, which are represented as hash values (cf. section 16.4), compares them to find intersections, and finally outputs all intersections found to the respective parties. The technical details of the Boolean circuit are presented in the Annex in section 16.9.5. This circuit, however, is not constant over the lifetime of MPC-based ExTRUST as it depends on the number of parties N (states can be added/removed) and inputs u (vulnerabilities can be added). The participating parties perform an initial MPC protocol prior to the actual execution to determine the maximum number of vulnerabilities among all parties, which then determines the number of inputs for the Boolean circuit that is evaluated by the MPC protocol. Now that every party knows the number of inputs to the Boolean circuit, each party in the fixed subset n of the N parties locally compiles the Boolean circuit that is inserted into the MPC protocol, *i.e.*, no further interaction is required by the parties to agree on the Boolean circuit. Malicious-secure MPC protocols ensure that parties, who compiled a fake Boolean circuit that does not compute the agreed functionality, are identified by the other parties.

A party cannot revoke or modify submitted vulnerabilities because they remain in the input list of the N actors. The parties can opt in and out by sending a notification message to the N servers. Only the inputs of the participating parties that are logged in are taken into account for the computation.

Complexity and optimization of the Boolean circuit

For N parties, state-of-the-art MPC protocols require sending and receiving $\mathcal{O}N$ messages for each `AND` gate in the Boolean circuit (Beaver et al., 1990), while `XOR` gates can be computed locally without any interaction between the parties (Lindell et al., 2016). Consequently, we optimize the number of `AND` gates in our Boolean circuit that is evaluated via MPC. In order to prevent the concrete set sizes of the individual parties from being leaked, we specify an upper limit u that determines how many inputs a party feeds into the circuit. If a party has fewer than u inputs, it fills the missing inputs

with random dummy values, which will not represent any vulnerability and thus will not occur in any intersection as the probability that two parties independently choose the same random dummy values is negligible. On a high level, every party inputs two unique keys – k_0 and k_1 – for each of its vulnerabilities into the Boolean circuit. The Boolean circuit outputs k_1 if this vulnerability is part of an intersection or k_0 if only the respective party knows this vulnerability. Although the resulting keys are leaked to all parties, only the party who input the keys learns any information about the intersecting identifiers of their vulnerability. For details, we refer to section 16.9.5.

Using Private Set Intersection to calculate collisions

In order to calculate the intersection of different stockpiles, we use the Private Set Intersection (PSI) protocol. PSI allows two parties to securely compute the intersection of their private sets without leaking any information about set elements that are not part of the intersection to the other participating party.

Multi-party PSI (Hazay & Venkitasubramaniam, 2017) extends the PSI functionality to more than two parties, *i.e.*, the parties jointly compute the overall intersection of all their input sets without leaking any information of set elements that are not included in the intersection. Unfortunately, multi-party PSI only outputs the set intersection of *all* input sets. However, in our exploit depletion system we search for intersections between *at least* two sets. A possible solution to this is to implement two-party PSI protocols between each pair of parties. However, this would require a quadratic number of protocol runs in the number of parties. Even more critically, this approach would reveal *which* other party has a common vulnerability. Instead, we use a generic MPC-based approach for our MPC-based ExTRUST application that is based on Huang *et al.*'s (Huang et al., 2012) Boolean circuit for two-party PSI which we extended into a multiple parties variant. A detailed description can be found in section 16.9.5.

Instantiation

There are many MPC frameworks based on secret sharing and/or garbled circuits, *e.g.*, (Aly et al., 2018; Bogdanov et al., 2008; Chaudhari et al., 2019, 2020; Demmler et al., 2015; Keller, 2020; Mohassel & Rindal, 2018; Patra & Suresh, 2020). Section 16.1 lists and compares several MPC frameworks with malicious security.

Since untrusted actors deal with highly sensitive information, we need security against malicious parties actively manipulating the computation to either learn more information or prevent other parties from receiving the correct output.

Current MPC frameworks that meet these requirements are MP-SPDZ (Keller, 2020) and SCALE-MAMBA (Aly et al., 2018). We recommend the use of *MP-SPDZ*, which implements, among other protocols, the constant-round BMR protocol (Beaver et al., 1990), which has benefits over the multi-round protocols of SCALE-MAMBA in high-latency networks. BMR is secure against malicious parties and a dishonest majority (*i.e.*,

Framework	# Parties N	Threshold t
ABY^3 (Mohassel & Rindal, 2018)	3	1
Sharemind (Bogdanov et al., 2008)	3	1
ASTRA (Chaudhari et al., 2019)	3	1
BLAZE (Patra & Suresh, 2020)	3	1
Trident (Chaudhari et al., 2020)	4	1
MOTION (Braun et al., 2022)	≥ 2	1
SCALE-MAMBA (Aly et al., 2018)	≥ 2	$N-1$
MP-SPDZ (Keller, 2020)	≥ 2	$N-1$

Table 16.1: Comparison of MPC frameworks that are secure against malicious adversaries, compute on Boolean circuits and allow up to t corruptions.

# Vulnerabilities \ # States	2	5	10	15
100	2	14	62	146
500	4	31	134	314
1000	7	49	210	492

Table 16.2: Runtime in minutes of MPC-based ExTRUST for various numbers of maximum vulnerabilities and states.

up to $N-1$ parties can be corrupted). If the number of computing servers is fixed to N=3 one can use ABY3 (Mohassel & Rindal, 2018), Sharemind (Bogdanov et al., 2008), ASTRA (Chaudhari et al., 2019), or BLAZE (Patra & Suresh, 2020); if the number is fixed to N=4 parties, Trident (Chaudhari et al., 2020) can be utilized. The MOTION framework (Braun et al., 2022) allows MPC among any number of parties N, however, it does not fulfil the full-threshold requirement of t=$N-1$. Table 16.1 shows an overview of the mentioned MPC frameworks.

16.6.2 *Feasibility of MPC-based* ExTRUST *implementation*

MPC-based ExTRUST completely relies on the security properties of the underlying multi-party computation (MPC) framework. While most MPC frameworks are implemented for academia usage, Bosch developed *Carbyne Stack* an open-source cloud stack for scalable MPC applications (Bosch Global, 2021) that is also suited for real-world usage. As the name suggests, the long-term plan is to make this MPC framework scalable for many participating parties. As this entire project is open-source, a group of states can use their implementation as basis MPC-based ExTRUST.

16.6.3 *Evaluation of the scalability of MPC-based* ExTRUST

In this section, we estimate the feasibility and scalability of our MPC-based ExTRUST. Since we know the complexity of our Boolean circuit, we can estimate the scalability of MPC-based ExTRUST.

In a realistic setting of our proposed application context of arms control, we have the following parameters for our benchmarks in section 16.2: number of parties / states $N \in \{2, 5, 10, 15\}$, maximum number of inputs $u \in \{500, 1000, 1500\}$, and length of vulnerability identifier hashes σ 256 bit. With these parameters, the size of our Boolean circuit is $\approx 4.8 \cdot 10^7$ `AND`s.

To estimate the runtime of our system, we generate a random circuit with the same number of `AND` gates and two `XOR` gates per `AND` gate. Since `XOR` gates can be evaluated in the BMR protocol without any communication (Lindell et al., 2016), it is less important to determine the exact number of `XOR` gates, as communication is the bottleneck of MPC.

For malicious MPC with a dishonest majority, as required by our adversary model presented in section 16.3.2, we use the constant-round BMR protocol (Beaver et al., 1990) using the MASCOT protocol (Keller et al., 2016) to compute the garbled tables as implemented in the MP-SPDZ framework (Keller, 2020). To conduct our experiments, we use five servers, each equipped with an Intel Core i9 processor with 2.8 GHz and 128 GB DDR4-RAM. The round-trip network latency in our simulated WAN setting is about 100 ms and the bandwidth 90 Mbit/sec. We take the average runtime of three executions.

The execution time of our circuit is about 31 minutes. This is an acceptable runtime for governmental actors, as the protocol is run daily or weekly. However, the size of the Boolean circuit and the cost of computing each `AND` gate are quadratic in the number of servers N. Therefore, our scheme will not scale for a large number of parties $N \gg 10$.

We can improve the scalability for these scenarios by outsourcing the computation to $n \ll N$ non-colluding servers (Kamara et al., 2011). Here, the N parties distribute their input to the n servers, which together run the MPC protocol and distribute the result. An advantage of this method is that all $N \gg n$ parties may be malicious as long as they can trust that the n servers are not colluding. This improves the cost of computing an `AND` gate to On^2.

16.7 Discussion

This section will discuss our approach. As the main contribution of this paper is the MPC-based ExTRUST, the BC-based ExTRUST prototype is not covered here, as it was discussed in section 16.5.3. In the following, we will analyze our MPC-based ExTRUST regarding the conceptual requirements RC1 - RC5 (section 16.7.1) and the security requirements RS1 - RS3 (section 16.7.2) necessary to create incentives for states to participate. An overview of which conceptual and security requirements are fulfilled by MPC-based ExTRUST and BC-based ExTRUST, respectively, is provided in section 16.3. This section also reviews the scenarios in which ExTRUST can be of use, followed by an outlook on possible future applications (section 16.7.3).

Requirement	BC-based ExTRUST	MPC-based ExTRUST
RC1	(✓)	✓
RC2	✓	✗
RC3	✓	✓
RC4	✓	✓
RC5	✓	✓
RS1	✗	✓
RS2	(✓)	✓
RS3	✓	✓

Table 16.3: Comparison of which conceptual RC1 - RC5(cf. section 16.3.1) and security requirements RS1 - RS3 (cf. section 16.3.3) are fulfilled by BC-based ExTRUST̃(cf. section 16.5) and MPC-based ExTRUST (cf. section 16.6). The bracketed checkmarks highlight requirements that are only fulfilled if we can exclude the 51% attack against Blockchains.

16.7.1 *Conceptual Requirements*

The MPC-based ExTRUST architecture allows participating parties to input information about their known vulnerabilities and exploits without openly revealing sensitive information to other parties (RC1). The output of the computation is the matching vulnerabilities and exploits between the parties (RC3). Since the output is computed interactively between the parties, MPC-based ExTRUST is entirely decentralized and does not require a trusted third party (RC5).

The solution is theoretically scalable to N=10 parties (RC4). However, the more parties are involved in the protocol, the more inputs and data have to be exchanged between these parties, *i.e.*, the approach has a complexity $O(N^2)$, but usually $N \approx 10$ (cf. section 16.3.1). However, as explained in RC4, real time computability is not a critical requirement and longer computation times are no problem for such highly politically organized processes like arms control, which often require days or weeks for the full formal process and the involvement of all necessary stakeholders. In section 16.6.3, we propose to outsource (Kamara et al., 2011) the computation to $n \ll N$ parties, which improves the performance of MPC-based ExTRUST. Considering the context of arms control, such a scenario is only applicable and likely if the outsourced computation is performed by neutral institutions that are not involved in the arms control measure itself, since in this way none of the parties involved need to trust that the other participants will not share information outside the protocol. Such delegation is not uncommon for arms control measures. An example is the Joint Comprehensive Plan of Action (JCPOA), a multilateral treaty known as the Iran nuclear deal (Maslov & Ustinov, 2020) between Iran, China, France, Russia, the United Kingdom, and Germany. The International Atomic Energy Agency (IAEA) manages and organizes all aspects of this treaty via independent bureaus, entrusted laboratories, UN working groups, and neutral experts for investigation field trips. Regardless, for practical arms control measures as our proposed depletion system, the amount of involved parties usually does not exceed a single-digit number and is often established between a small group of states.

Unfortunately, requirement RC2 is not met because intersection checks now require interaction, as the participating parties are required to exchange data. However, exactly this property of ExTRUST is the key to avoid local brute-force attacks, to which BC-based ExTRUST is vulnerable (cf. section 16.5). Although, in the context of arms control, the minimum threshold for cooperation to which states must commit provides an incentive to join the measure, this requirement is not mandatory to practically operate ExTRUST. As a privacy-preserving arms control measure is more critical than the desire to independently check for intersections, we consider this a weak limitation that does not undermine the practical value of our approach, especially when considering that ExTRUST fulfills all other conceptual requirements.

16.7.2 *Security Requirements*

MPC-based ExTRUST Fulfills all three security requirements presented in section 16.3.3. A notable advantage of MPC-based protocols is that the participating parties can only derive information from their own inputs and the outputs received, *i.e.*, the parties do not learn more information in the MPC-based ExTRUST than in ExTRUST with a trusted third party that receives the inputs from all parties and outputs the intersections. This means that no more information is revealed in the protocol transcript than an adversary would learn in the ideal world. In contrast to BC-based ExTRUST from section 16.5, local attacks (*e.g.*, brute-forcing specific hash values) are not possible in MPC-based ExTRUST. In addition, an adversary is not able to copy vulnerabilities or exploits from other parties to output an invalid intersection because the inputs are inaccessible to the other parties. Thus, requirement RS1 is completely fulfilled.

Once the inputs are submitted in the MPC protocol, the parties are not able to withdraw or modify them (RS2). A situation in which all state actors jointly manipulate the protocol will never happen, since they could otherwise share their vulnerabilities in plain anyway.

False positive intersection (RS3) results are possible with a negligible probability. A false positive is possible if two different vulnerabilities are mapped to the same hash value. Since we use a collision-resistant hash function, the probability of other collision scenarios is negligible. In addition, a false positive may occur if two parties independently choose the same key identifiers. Due to the usage of 256 bit key identifiers and a robust random generator, the probability of this situation is negligible as well.

Above all, the security and confidentiality of the assets to be shared are key incentives for establishing an arms control measure. As MPC-based ExTRUST fulfills all security requirements, it is suitable for a real world application without discouraging states from using it.

16.7.3 *Further Application Scenarios*

Beside the proposed context, ExTRUST can also be useful in other application scenarios, some of which will be discussed in the following.

At present, our approach concentrates exclusively on state actors as addressees. However, organizations or individuals might also be interested in using such a system. As explained in section 16.2.2, bug-bounty programs and vulnerability research projects have similar goals: to reduce the spread of vulnerabilities to secure systems. Here, using the aggregated information from the external stockpile depletion measures and integrating it into ExTRUST can increase the speed of detection of matching rediscoveries in stockpiles. This can be achieved by using writers for selected public services or other institutions that intend to contribute to cybersecurity, which feed their hashed vulnerability identifiers into ExTRUST. In such a setting, the hashing of information is as important as in ExTRUST's motivational scenario to prevent the material from being disseminated for malicious cyber operations. The use of ExTRUST in a purely corporate environment is probably not possible, as organizations like Zerodium (ZERODIUM, 2019) are primarily looking for exploits to sell. A similar bug-bounty related approach could focus on examining discovered, potential zero-day vulnerabilities against other submitted but not yet publicly disclosed vulnerabilities. The history of submissions would allow submitting actors to claim their first-submitted-reward later on, once the information is disclosed. In this way, the first finder could be paid out without the hackers having to reveal their discovery in advance.

16.8 Conclusion and Future Work

The paper focused on the depletion of vulnerabilities and exploits that are being stockpiled by state actors, a development that is accelerated by enhancements in the field of AI and its military application. While the disclosure of vulnerabilities at the national level through regulatory processes is becoming more and more of an issue, cooperation on disclosure at the bilateral or multilateral level is still lacking. We discussed that an important obstacle to such measures is the comprehensible restraint of states to give up their accumulated intelligence information in order to compare stockpiles and unnecessarily reveal unique exploits or other secret assets.

In order to develop a technical measure in such a zero-trust scenario, we identified structural as well as IT security requirements for the detection of intersections in different exploit stockpiles. Based on these, we discussed and designed (i) a novel identification scheme for vulnerabilities and exploits and (ii) an external, privacy-preserving exploit depletion system named ExTRUST.

We have identified the requirements for this depletion system for zero-trust relationships and shown that the technical security requirements could hamper the political incentives for states to cooperate. We have illustrated this challenge by developing a prototype for a depletion system based on a Blockchain. The presented MPC-based ExTRUST system handles this dualism by focusing on the IT security of a depletion system while fulfilling most of the conceptual requirements. It stores the detected intersections, while the submitted vulnerabilities are protected by the MPC protocol and thus remain hidden from all involved actors. However, one limitation of this approach is that it is vulnerable to collusion by multiple actors, as they could add vulnerabilities and remove them from the intersection – an edge case that is not an option in the proposed arms control context. We also argued that MPC-based ExTRUST is currently not able to fulfil all conceptual

requirements, as participating states need to explicitly cooperate and share obfuscated information, which could be a disadvantage regarding its implementation. Nevertheless, we have shown that the strength of the MPC protocol lies in the fact that an adversary cannot obtain more information from the joint computation than if a trusted third party were to compute the intersections. The ExTRUST system uses a novel exploit identifier and discussed how this identifier could be improved in different scenarios to address the trade-off between the uniqueness and ambiguity of the properties. We believe that this provides a secure measure which fulfills the state's need for secrecy and yet at the same time can contribute to the reduction of vulnerability stockpiles in order to foster the public IT security through the disclosure of vulnerabilities.

We discussed further application scenarios beyond the specific context of cyber arms control with different parties comparing their vulnerability stockpiles. We demonstrated that such an approach could be facilitated for external depletion measures such as bug-bounty programs. Such measures could potentially be extended so that even private actors could contribute to the internal exploit stockpile depletion process by adding external information about the depletion into ExTRUST.

As the discussion has shown, further evaluation and study of our concept is recommended, in particular in terms of the definition of the identifier. We discussed that a current limitation of the identifier is the necessity to find a sweet spot in the accuracy regarding the description of a security vulnerability that prevents duplicate descriptions of the same identifier while avoiding an unnecessary and potentially problematic generalization.

Future work should analyze the relationship between the uniqueness and ambiguity of the characteristics of the identifier, the size of the identifier space, and – on a practical level – whether security experts independently create matching identifiers for the same vulnerability. Further work should focus on the possibility, the role and the security requirements of a trusted third party like the UN to calculate stockpile intersections, in order to circumvent the current necessity of cooperation between potentially opposing state actors. In addition, it would be interesting to implement ExTRUSTas an actual measure between state parties to monitor its real world usage, its perception of the systems security and usability by the participating states as well its impact on their vulnerability disclosure considerations.

Due to the high political relevance of our proposal, we hope that this approach can be an inspiration to computer science and engineering to reflect on the ethical responsibility for the domain of cyberspace and its peaceful development and that future interdisciplinary work in this area will bring together researchers from privacy, IT security, and peace and conflict research.

16.9 Annex

16.9.1 *Required properties for a machine-readable vulnerability identifier*

In order to automatically compare vulnerabilities, our approach requires a vulnerability identifier $id\left(i_v^p\right)$ for a given vulnerability v, which must be unique among all parties $p \in \mathbb{P}$ sharing v's information i_v. Hence, the equation $id\left(i_v^p\right) = id\left(i_v^{p'}\right)$ should hold for all $v \in \mathbb{V}$ and $p, p' \in \mathbb{P}$, where i_v^p and $i_v^{p'}$ are the vulnerability v's information of the respective party.

16.9.2 *Ambiguous vulnerability identifier*

The vulnerability identifier description is ambiguous, if it is possible to describe two different vulnerabilities $v_1, v_2 \in \mathbb{V}$, where $v_1 \neq v_2$ with the same identifier, *i.e.*, $id\left(i_{v_1}^p\right) = id\left(i_{v_2}^p\right)$, or to use two different identifiers for the same vulnerability $v \in \mathbb{V}$, *i.e.*, $id\left(i_v^p\right) \neq id\left(i_v^{p'}\right)$.

16.9.3 *Approximation of the vulnerability identifier space*

Given the number of CWE classes, the size of the CPE directory and space of possible function names *FN* the vulnerability identifier space can approximated to $|id| = |CPE| * |CWE| * |FN|$, with $|CPE| \approx 2^{19}$, $|CWE| \approx 2^9$, and $|FN| \approx 2^{53}$ (using the NLTK English word corpus) resulting in $|id| \approx 2^{81}$.

16.9.4 *The* `checkIntersections` *transaction of BC-based* ExTRUST

```
/**
 * Sample transaction processor function.
 * @param {de.peasec.vulnerability.CheckCollisions} tx Sample transaction instance
 * @transaction
 */
async function checkCollisions(tx) {
	const assetRegistry = await getAssetRegistry('de.peasec.vulnerability.
	    ↪ Exploit');
	const allExploits = await assetRegistry.getAll();
	const collidingExploits = {};

	for (const exploit of allExploits) {
		const id = exploit.id;
		const hash = exploit.vulnerabilityIdentifierHash;

		if (!(hash in collidingExploits)) {
			collidingExploits[hash] = [exploit];
		} else {
```

```
                    collidingExploits[hash].push(exploit);
                }
            }

            for (const property in collidingExploits) {
                if (collidingExploits.hasOwnProperty(property) && collidingExploits[
                    ↪ property].length >= 2) {
                    exploits = collidingExploits[property];

                    let event = getFactory().newEvent('de.peasec.vulnerability',
                        ↪ 'CollisionFound');
                    event.affectedExploitIds = exploits;
                    emit(event);
                }
            }

}
```

Listing 16.2: The function that checks for matching items in the BC-based ExTRUST ledger.

16.9.5 *PSI-variant Boolean circuit for multiple parties*

Our circuit extends Huang *et al.*'s (Huang et al., 2012) Boolean circuit for two-party PSI of complexity $O(u \log u)$ into a PSI-variant circuit for multiple parties with special-purpose filter options for matching items (cf. section 16.9.5). In the first step, each party i locally sorts (*i.e.*, outside the circuit) its set of triples $X_i = \{x_{i,1}, \ldots, x_{i,u}\}$ s.t. $v_{i,j} < v_{i,l}$ if $j < l$ and inputs this into the circuit. The first task of the circuit is to verify that the input set of each party is correctly sorted. This is a linear sweep through the input set of each party with u comparison circuits for the respective $v_{i,j}$-values and has total size of $O(Nu\sigma)$ `AND`s (Kolesnikov et al., 2009). We open a flag l_i for each party i to all parties by using reactive MPC, that indicates if i's inputs were correctly sorted. If one of the parties cheats ($l_i = 1$), the protocol aborts and the cheating parties are removed from future computations.

The next part of the circuit is similar to that of Huang *et al.* (Huang et al., 2012): the single sorted sets are oblivious merged to one large sorted set. For this purpose, we span a binary tree of sorted lists, where each node on depth i implements a bitonic merger (Huang et al., 2012) for $2^i u$ inputs, which has a complexity of $O\left(2^i u\sigma \log\left(2^i u\right)\right)$ `AND`s. A bitonic merger takes two sorted sets as input and outputs the merged sorted set. In total, we have $N-1$ of these bitonic mergers and the binary tree has a depth of $\log_2(N)$. This results in an upper limit of $\sim 2N^2 u\sigma \log(Nu)$ `AND`s. In our special case, all triples $x_{i,j}$ are now sorted according to the values $v_{i,j}$ of $x_{i,j}$, so that matching vulnerabilities lie directly next to each other. Since the triple can no longer be assigned to a specific party, we denote the outcome of this subcircuit with $x_1, \ldots, x_{uN}$.

Afterward, neighboring entries are compared with a *DupKeySelect* block. A DupKeySelect block takes as input two neighboring triples $\left(v_i, k_i^0, k_i^1\right)$ and $\left(v_{i1}, k_{i1}^0, k_{i1}^1\right)$, and outputs $\left(k_i^1, k_{i1}^1\right)$ if $v_i = v_{i1}$ or $\left(k_i^0, k_{i1}^0\right)$ otherwise. We compare the first and last value with a zero string 0^σ to avoid leaking information about the frequency of occurrence of a key. The corresponding circuit a has size of $\sim Nu\sigma$ `AND`s.

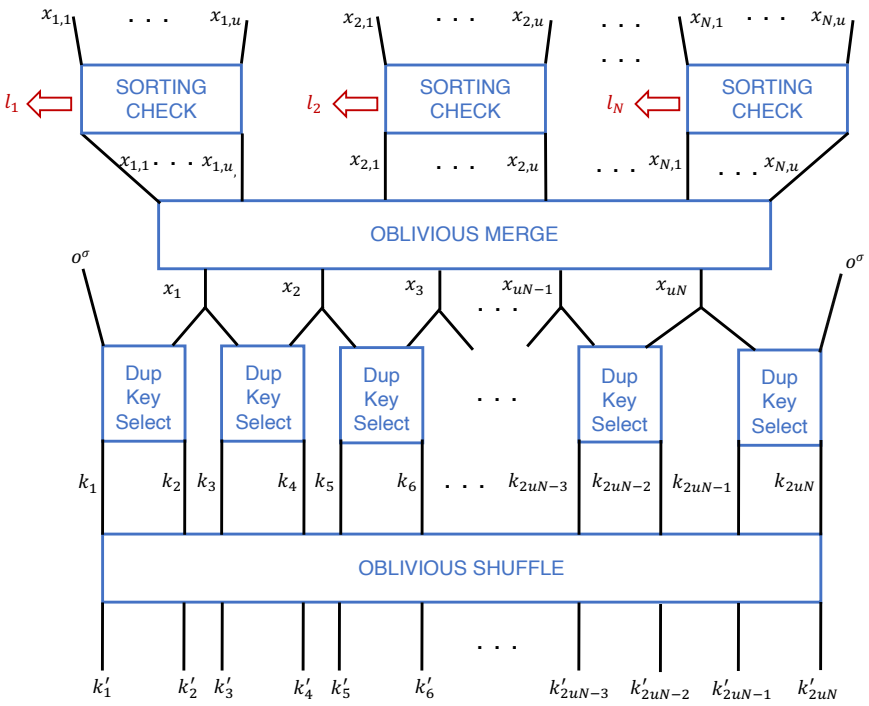

Figure 16.3: Boolean Circuit design output identifier keys that are dependent if their respective vulnerability identifier occurs at least twice. The complexity of the circuit is $\sim 2Nu\ 2N^2u\sigma \log(Nu)$ ANDs for number of parties N, number of inputs per party u and security parameter σ. The design is based on Huang *et al.*'s Boolean circuit for computing the set intersection between two parties (Huang et al., 2012). The red values $l_1, \ldots, l_n$ are opened and verified before the oblivious merge block is executed.

The final step is to shuffle the output keys $k_1, \cdots, k_{2uN}$. This circuit has a complexity of $O(Nu\sigma \log(Nu))$ ANDs (Waksman, 1968). The resulting keys $k'_1, \ldots, k'_{2uN}$ are opened to all parties. Finally, the parties can identify matching vulnerabilities by checking which of their input keys occur in the output. Overall, the complexity of the circuit is clearly dominated by the oblivious merge part of size $\sim 2N^2u\sigma \log(Nu)$ ANDs.

During the lifetime of ExTRUST, the parties cannot update the hash values $v_{i,j}$ from $x_{i,j}$, but they use fresh identifier keys $k^0_{i,j}$ and $k^1_{i,j}$ in each protocol run.

Generalization of MPC-based ExTRUST

The above described circuit implements a variant of the private multi-party intersection, where an intersection is found when at least two parties share an element. However, this circuit can easily be adapted to other variants of the functionality relevant for arms control. We provide two examples below and note that adapting to other variants is a simple circuit design task.

At least m parties In some cases, it may be useful to know whether elements are shared by at least m parties (say m 3) in order to decide whether a vulnerability or exploit can be made public. This can be easily achieved by replacing the *DupKeySelect* block in section 16.3 with an *m-DupKeySelect* block, which outputs the 1 keys if m neighboring elements are equal. Such an *m-DupKeySelect* block has a complexity of $O(m\sigma)$ ANDs so that the total complexity of this subcircuit is $O(mNu\sigma)$ ANDs which is still negligible compared to the rest of the circuit.

At least z fixed parties and m other parties In this scenario, the parties aim to find intersections that z fixed parties know about (*e.g.*, China and USA with z 2) and at least m other parties. This can be achieved by adding a fixed identifier t to the input tuples of the parties (*e.g.*, set to 0 for China, 1 for the USA, 2 for all others). The *DupKeySelect* block in section 16.3 is then replaced by a *Programmable-*$(m\ z)$*-DupKeySelect* block, which receives a programming bit p as additional input, indicating whether a duplicate check is valid or not. More specifically, if p 1, the block will output the result of an $(m\ z)$*-DupKeySelect* block, and otherwise, if p 0, the block will never output an intersection. The programming bit for these blocks is indicated by a *z-Filter* block, which takes the parties' identifier t and checks if the z fixed parties are part of the potential intersection (*e.g.*, $t \in \{0, 1\}$).

The *z-Filter* block takes as input z m identifiers $t_1, \ldots, t_{zm}$, has a set of fixed identifiers $T_1, \ldots, T_z$, outputs one single bit p and works as follows: We check for all fixed identifiers T_i if an input identifier t_j exists where $T_i = t_j$ holds. We do this by comparing T_i to all input identifiers and using OR-gates to fold the result to a single bit. The total complexity of this step for all fixed identifiers is $O(z\omega(m\ z))$ ANDs, where ω $\log(z\ 1)$ is the bit-length of the state identifiers. Afterward, we use $z - 1$ AND gates to fold the z resulting bits to one bit p, which results in a total complexity of $O(z\omega(m\ z))$ ANDs.

The *Programmable-*$(m\ z)$*-DupKeySelect* block takes $z\ m$ quadruples $x_i = \left(v_i, k_i^0, k_i^1, t_i\right)$ as input and outputs $z\ m$ keys $k_1^{0/1}, \ldots, k_{zm}^{0/1}$. It consists of a $(z\ m)\,\sigma$-bit multiplexer (one bit for each output bit), where its first input consists of the input keys $k_0^0, \ldots, k_{zm}^0$, the second input is the output of the $(m\ z)$*-DupKeySelect* block, and the programming bit is the output of the *z-Filter* block p. The multiplexer circuit has a size of $O\left(\sigma\left(z\ m\right)\right)$ `AND`s. Overall, the complexity of the *Programmable-*$(m\ z)$*-DupKeySelect* block is dominated by the size of the $(m\ z)$*-DupKeySelect* block of size $O\left(mNu\sigma\right)$ `AND`s which is negligible compared to the rest of the circuit.

17

VERIFICATION IN CYBERSPACE

ABSTRACT Verification is one of the pillars of arms control and non-proliferation treaties, as well as an important part of confidence building measures. It defines practical measures that enable treaty members to control the treaty compliance by observing, counting or monitoring specific actions and their accordance with the respective rules. In contrast to historical examples of former military technologies, cyberspace features some unique characteristics, making it hard to apply established measures. The chapter describes these peculiarities and assesses distinguishing problems compared to selected established verification measures for nuclear, biological and chemicals weapons technology. Yet, cyberspace is a man-made domain and adjusting its technical setting, rules and principles may help to reduce the threat of ongoing militarization. Offering some alternatives, the chapter elaborates on suitable and measurable parameters for this domain and presents potentially useful verification approaches.

ORIGINAL PUBLICATION *Reinhold, T., & Reuter, C. (2019c, March 13). Verification in Cyberspace. In C. Reuter (Ed.),* Information Technology for Peace and Security: IT Applications and Infrastructures in Conflicts, Crises, War, and Peace *(pp. 257–275). Springer Vieweg. https://doi.org/10.1007/978-3-658-25652-4_12*

17.1 WHAT IS VERIFICATION?

The international law is based – among other things – on treaties and binding agreements between states that define the rules for state behavior and its interactions. One of the main principles of these rules is the convention "**pacta sunt servanda**" (Wehberg, 1959), which basically translates to "agreements must be kept". While the principle had been state practice for centuries, its first explicit reference was made in 1969 in the "Vienna Convention on the Law of Treaties" which describes that "every treaty in force is binding upon the parties to it and must be performed by them in good faith" (UN, 1969). Therefore, it highlighted the question which instance should be in charge of controlling the compliance of states to specific treaties and how this should be performed. This question had been answered over the last decades in different variations, led by the principle that states are sovereign entities and to a high degree autonomous in their decisions, which mainly rules out the possibility of "higher instances"[1]. Therefore, states basically regulate their relations by mutual agreements. A complementary tool for treaties is the possibility of treaty partners to control the compliance of each other by practical measures, the so-called **verification**. Verification often belongs to international

[1] State sovereignty is one of the core principles of the UN Charta, which defines the general rules of state behavior and interstate relations. Only dedicated institutions like to UN security council are authorized to restrict this right.

treaties but can also be part of non-binding interstate agreements in terms of confidence and trust building among opposing state actors[2] that thereby are able to demonstrate their good intent. Verification is an important measure for international security politics and mostly integrated in so-called **verification regimes**, a concept that is based on the regime theory of Robert O. Keohane (Keohane, 1984). His theory describes "*institutions possessing norms, decision rules, and procedures which facilitate a convergence of expectations"* (Krasner, 1983). In theory, a regime is a set of *"principles, norms, rules, and decision-making procedures around which actor expectations converge in a given issue-area*" (Krasner, 1983). In terms of verification, this means that a verification regime consists out of the following different parts that the affected states negotiated and agreed upon:

- The agreements itself.
- The specific thresholds, binding instructions or forbidden activities that belong to rules which the treaty members agree to follow.
- The practical measures that treaty members or specifically entrusted authorities are allowed to perform in order to control the compliance of the treaty members.
- Optionally the definition of the authority that is allowed to decide over the compliance and the consequences that states agree to perform and bear when the agreed rules are not followed.

Verification regimes had been developed over the last decades for different reasons and situations and are based on different mandates, often in the context of disarmament, arms control or the so-called **non-proliferation**[3] of military technology. Every regime is based and depended on the political acceptance of the agree measures. A popular example for verification in the context of nuclear armament is the **International Atomic Energy Agency** (IAEA), an independent international organization that reports to the United Nations General Assembly and the United Nations Security Council. With the international adoption of the Treaty on the **Non-Proliferation of Nuclear Weapons** (NPT)[4], the IAEA had been put into charge in different treaties (Neuneck, 2017) "*to establish and administer safeguards designed to ensure that special fissionable and other materials, services, equipment, facilities, and information made available by the Agency or at its request or under its supervision or control are not used in such a way as to further any military purpose; and to apply safeguards, at the request of the parties,*

[2]Confidence and trust building (CBM) is a measure to establish the cooperation of states by stepwise mutual concessions, information sharing and the reduction of military pressure. CBM as a concept had been developed by the Conference on Security and Co-operation in Europe (CSCE) during the Cold War era (Bazin, 2013).

[3]Proliferation is a concept from international security politics that describes the spread or the intensification of the knowledge, the technology or the material of a specific military weapons technology. It is further graduated in horizontal proliferation (the spread to new states that don't dispose this specific military technology) and vertical proliferation (the advancement and stockpiling of one state for a specific military technology). Non-Proliferation contains measures of arms control like treaties and agreements that should prevent this spreading.

[4]The Treaty on the Non-Proliferation of Nuclear Weapons ("Non-Proliferation Treaty", NPT) is an international treaty that entered into force 1970 and whose objective is to reduce and prevent the spread of nuclear weapons and their technology and instead foster the peaceful application of nuclear energy (NPT, 1970).

to any bilateral or multilateral arrangement, or at the request of a State, to any of that State's activities in the field of atomic energy" (IAEA, 1961).

One of its most recent tasks is to control the compliance of Iran to the JCPOA nuclear treaty agreements (Joint Comprehensive Plan of Action) (IAEA, 2016) that had been negotiated over the last years and came into force in January 2016. Verification measures are integrated as so-called **safeguards**. They enable IAEA staff members to get access to nuclear and research facilities, shut down and seal critical industrial hardware, install surveillance cameras, control industrial plants, count the equipment in nuclear facilities, take samples from nuclear material as well as measure the radiation level of devices and places. As already pointed out, these verification measures are always practical steps that tightly concentrate on specific aspects of the controlled technology and whose outcome can be compared against threshold values, "do's and don'ts" or lists of forbidden technological procedures. Another example of a verification regime concerns chemical weapons and feasible weapons material. This regime had been put in place according to the **Chemical Weapons Convention** (CWC)[5], an international arms control treaty that had been negotiated by the UN and entered into force in 1997. The treaty *"(..) prohibits the development, production, acquisition, retention, stockpiling, transfer and use of chemical weapons. It also prohibits all States Parties from engaging in military preparations to use chemical weapons"* (Boehme, 2008) and it is administered by the **Organization for the Prohibition of Chemical Weapons** (OPCW) which had been explicitly founded for the task of verification. All verification measures of the CWC are defined and signed by the treaty members in a dedicated "verification annex". This annex contains detailed explanations which and how verification measures are performed, lists the allowed measurement procedures, defines who is entitled to perform specific tasks and analyze the taken samples and how the results are reported (Boehme, 2008). A key element of the CWC are the inspections to control industrial plants as well as civil and military research facilities and laboratories, monitor the production of critical chemical materials, count fabrication materials and equipment, take chemical samples and control for specific forbidden military "delivery systems"[6].

In regard to former military developments, verification measures like the described examples had been put in place in situations when new technical advancements or innovations significantly destabilized the international balance of powers, led to arms races or contained the potential for massive destruction or unutterable suffering. In these situations, verification has been a measure to sustain and support political stabilization agreements by mutual control mechanisms. When looking at the current developments in the cyberspace, international security politics have to handle such a situation where military forces are quickly adopting to this new situation and consider the cyberspace the next military domain where defensive and offensive measures are necessary. More and more military forces are establishing dedicated cyber commands (UNIDIR, 2013) and alliances are fostering the establishment of collective capacities for military engagements. For example, the NATO decided in 2016 that the cyberspace is an essential domain that needs to get covered by the collective defense strategies and that attacks over the cyberspace can invoke the alliance case of Article V of the NATO Charta. This development raises many concerns due to the lack of international political regulation or even a common international understanding between states on how the rules of

[5]The full title of the treaty is "Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction"

[6]J. B. Tucker (Tucker, 1998) gives a comprehensive overview.

international law apply to this domain, what is allowed and what is prohibited. Although some suggestions had been made, like the so-called Tallinn Manual (CCDCOE, 2013) or the of Russia, China, Tajikistan and Uzbekistan (UN General Assembly, 2011) none of these approaches reached an international consent so far. This situation is tensed on the one hand by the fact that it is yet unclear how offensive tools for the cyberspace that can be targeted against IT systems (**cyberweapons**) can be classified in terms of their destruction potential and how this impact can be estimated. On the other hand, IT systems are an essential part of most societies and, due to their interconnected nature, critical for the global economy, a fact that is accommodated in many countries by the classification of IT systems and its networking hardware as critical infrastructure (for example, see EU-Parliament, 2008). With regard to the technical know-how of IT systems, the knowledge as well as the global economic players are concentrated in just a few countries that currently dominate this field of technology and therefore to a high extent also its military application. This led to a situation where it is rational for military decision makers and politicians to consider their countries as threatened by such military and potential destructive powers and to establish own military programs to counter this situation and keep the pace. An arms race has started.

17.2 THE SPECIAL CHARACTERISTICS OF THE CYBERSPACE DOMAIN

The described situation underlines the necessity of regimes for the cyberspace and related arms control measures to limit this development, establish binding rules and create a calculable situation for interstate relations. On the other hand, as has been already pointed out, this situation is rarely new, and states have faced similar circumstances over the last decades for other technological developments. It is therefore appropriate to gather insights from the former "lessons learned" and project them to the current situation. Unfortunately, this approach quickly reveals that the cyberspace has some unique technical specifics and features that differ strongly from other technical developments. These features, that will get briefly analyzed in the next part, hinder the projection of established arms control and verification measures to the cyberspace, and therefore have to be taken into account for the development of applicable measures.

17.2.1 *The Problems of Counting Data in a Virtual, Distributed Space*

The cyberspace is by design a "virtual" domain that abstracts a space from a specific real geographic location. It consists of autonomous, self-contained networks that integrate and connect groups of different IT systems, while each network itself can consist of smaller sub-networks. Any kind of data is on the one hand theoretically stored and processed by a specific IT system which has a geographical location and falls under a jurisdictionally responsible legislation. On the other hand, especially in times of the so-called **cloud computing**, data can seamlessly transferred to, copied to and stored in another system for availability or split up to multiple parts to be stored and processed on multiple, distributed IT systems. In either case, data itself has no specific physical

representation[7] that can be monitored and can be seamlessly duplicated. This situation makes the geographical pinpointing of a specific piece of data problematical and renders two main concepts of established verification meaningless: the counting and verifiable limiting of objects. Digital data does not produce any kind of reliable "traces" that might be used to monitor the actions of a specific institution or actor. This situation is furthermore complicated by the so-called **attribution problem** that – in a nutshell – describes the problems and the ambiguity of assigning any kind of activity within the cyberspace to its origin and the presumed actor that intentionally performed this activity[8].

17.2.2 *Dual Use: Technology for Civilian Purposes and Military Applications*

Another feature of the cyberspace, and especially of the technical equipment that is necessary for its infrastructure is its so-called **dual-use character**. The term describes the feature of specific goods[9] that can be used for military as well as civilian purposes without being able to draw a distinct line between these usage scenarios and which therefore cannot be generically prohibited for arms control reasons. Such goods need to get monitored in detail, because only their precise usage decides whether it affects negotiated agreements or not. Popular examples for dual-use goods are biological agents or other basic material for vaccines that are necessary for civilian health care reasons and medical research but can be also used for military purposes. The task of defining lists of such goods and its necessary special verification treatments into treaties is performed since several decades for nuclear, chemical and biological goods. Its most popular example is the Wassenaar Arrangement ("The Wassenaar Arrangement on export controls for conventional arms and dual-use goods and technologies - List of dual-use goods and technologies and munitions list") (The Wassenaar Arrangement Secretariat, 2022), a binding regime between currently 42 participating states that agreed upon sharing trade data of such sensitive goods as a measure of trust and confidence building as well as establish national export controls. The agreement had been extended in 2013 to cover so-called **intrusion software**, that is "specially designed or modified to avoid detection by 'monitoring tools', or to defeat 'protective countermeasures', of a computer or network capable device" ("The Wassenaar Arrangement on export controls for conventional arms and dual-use goods and technologies - List of dual-use goods and technologies and munitions list," 2017) (The Wassenaar Arrangement Secretariat, 2022) and able to either retrieve data from IT systems or alter their standard behavior. Nevertheless, the dual-use character of IT hardware and software is distinct and many argued that the new regulations of this extension could lead to problems with legitimate

[7] Of course, all pieces of data are stored physically in different ways (like magnetic fields and classic hard drives or electromagnetic states on solid state drives) but this stored data cannot be handled as an unique and autonomous, self contained entity like a missile, a tank or an test cube.

[8] The necessity of attributing an attack to its origin is a key element to the states right for self defense under the UN Charta. Nevertheless, attribution in cyberspace is hindered by multiple possibilities of adversaries to cover tracks and use IT systems of uninvolved third parties. Attribution cyberattacks is therefore currently considered to be the main problem when applying international law and its rules of state behavior to the cyberspace. As an example, see Guerrero-Saade and Raiu (Guerrero-Saade & Raiu, 2017).

[9] The term "goods" which includes software as well as technology is used especially in dual-use scenarios of arms control and non-proliferation to describe "anything that needs to get regulated" without being exclusively restricted to military technology and with explicit inclusion of necessary base materials for potential military products.

research on cybersecurity measures if restrictively put into force (for example, see Hinck, 2018). Compared to former dual-uses approaches, a relevant factor for national trade regulations of chemical, biological or nuclear goods was either the sheer amount of specific materials, the necessary equipment or specific military delivery systems that can be controlled. For the cyberspace this is not possible, because both the hard- and software and their extent are the same for civil, economic and military purposes. 2018).

17.2.3 *Differentiation between Defense and Offense*

A last aspect that is strongly connected to the dual use debate is the differentiation between goods that distinctively serve for military defensive measures and those whose primary purpose is for offensive measures. Such differentiation could be deployed for the regulation and application of verification measures on the trade, possession and usage of respective cyber capacities. Nevertheless, as pointed out before, there is no obvious distinction for IT goods due to their dual-use character. Even apparently offensive tools like malware or software exploits are necessary to test and increase the cybersecurity of one's own IT systems. A popular example for this case are so-called **penetration testing tools**, i.e. software that is specifically designed to attack and penetrate IT systems and networks to detect flaws, weaknesses and security problems. These tools are an important instrument for IT security practitioners and their regulation can affect the protection of IT systems. On the other hand, their detection during potential inspections therefore does not necessarily prove any non-compliance. An exception could be seen in "hand-crafted" software that is produced and dedicated solely for cyberattacks. It is supposed that such things might become more relevant in the next years when the economy increasingly adapts to the demand from military forces for such things. Nevertheless, the absolute majority of cyberattacks in the last years, even those with presumed state actors, had been carried out with "off the shelves" tools and software, which, due to the nature of rapidly changing technology in the cyberspace, is often the more effective way to perform the goals (as an example, see the 2018 version of Verizons "Data Breach Investigations Report" (Verizon, 2018)).

17.3 Established Verification Measures and their Problems when Applied to the Cyberspace

The previous glimpse on the established verification measures of other technological developments in the light of the technical specialties of the cyberspace already predict that applying or projecting these measures directly will certainly not work for this new domain (Pawlak, 2016). Nevertheless, to understand how applicable verification measures for the cyberspace can be developed, which problems arise and how they need to be differentiated from former approaches, it is helpful to understand the core principles of the established verification regimes and its control measures.

As has been pointed out, verification measures always control the compliance to agreements and although the previous examples illustrated that they strongly differ for various kinds of situations, all of them basically contain some of the following four restrictions and its adjacent verification principles (Neuneck, 2017):

- Geographical restrictions that regulate the allowed or prohibited location of specific goods which are controlled by locating and visually monitoring these goods (this might include ultraviolet and x-ray imaging as well as aerial and satellite photography)

- Limitations in terms of the overall amount or even the complete prohibition of the possession of goods are controlled by counting and cataloging the goods

- Definitions of threshold values for specific properties of physical, chemical or biological states of goods and military systems can be controlled by measuring or scientifically estimating these properties of goods

- Restricting the proliferation of goods that is controlled by regulating their trade and tracing the exported goods

With the technical specifics of the cyberspace in mind, it becomes clear that most of the established verification measures will not work for the cyberspace because their core principles are designed for physical domains like sea, air, land or space and rely on features of these domains that the cyberspace does not provide. This problem should be analyzed in detail.

The principles of geographical restrictions are undermined by the virtuality of the cyberspace. Even if hardware itself always has a physical representation, the storage and processing of data cannot be reasonably attributed to a geographical location. Also, where hardware can get monitored and controlled, it is not the hardware but merely the software and its usage that differs between legitimate usage and a theoretically forbidden application, a differentiation that is hard to make due to the dual-use character. Furthermore, even if one assumes the existence of specific military grade software, it is hardly practical to control IT systems regarding their installed software to search for theoretically forbidden offensive tools. IT systems provide numerous ways to hide data, e.g. so-called **hidden volumes** (Bellare & Rogaway, 2005), a cryptographic way to hide software or data within the apparently "free space" on storage devices that can only be detected and unlocked by insiders with specific software and passwords.

Controlling and tracing the proliferation of software and hardware is another principle that is rendered nearly impossible by its dual-use character. We are practically unable to decide whether they are used in a legitimate way. On the other hand, the virtuality of the domain cyberspace allows adversaries to cover their tracks or manipulate them to put investigators off the scent. The ongoing debates on the problems of attributing cyberattacks illustrate these problems in detail (as an example, see Guerrero-Saade and Raiu, 2017). Also, as pointed out before, only the usage of tools decides about the offensive or defensive application of goods, so any rules of verification regimes that declare forbidden behavior need to implement measures of controlling the specific application of IT goods, which is not practically implementable.

One principle where the cyberspace especially differs from other domains is the lack of a physical representation and, on the other hand, the seamless duplication of data. As argued before, malware and data cannot be counted – which might be a commonplace but renders any approaches of limiting specific things useless. For devices like IT

hardware that theoretically can be counted, the strong dual-use character again interferes with this approach of regulation[10].

The principle that seems to be most suitable to be projected to the cyberspace is the definition of any kind of thresholds as port of verification regimes. This paradigmatically builds on the idea that it is not the presence but the extent of the usage of goods that defines compliance or non-compliance, which strongly applies to the cyberspace. The question therefore is what parameters can be measured for the cyberspace and its underlying IT infrastructure and how measurement and controlling approaches can work.

17.4 APPROACHES OF VERIFICATION FOR THE CYBERSPACE

Despite the problems that had been pointed out in the previous chapters, verification for cyberspace has one strong advantage over other domains. In contrast to air, space, sea and land, the cyberspace is a completely man-made domain. Every rule and functional principle is defined and created by people or rather international committees like the standardization-focused **Internet Engineering Task Force** (IETF) (Bradner, 1999) or the more research-focused **Internet Research Task Force** (IRTF) (Sherry, 1996) that develop new technologies for the cyberspace and decide over their deployment. This means that – at least in theory – these principles can be adapted and further developed to support the peaceful development of this domain, to create transparency where it is necessary and support the establishment of measures for international political stability. Furthermore, the following sections will show that some necessary technical solutions that might be applicable for verification tasks already exist in other contexts of IT tasks.

17.4.1 *Measurable Parameters of the Cyberspace*

The question is which parameters of the cyberspace, its infrastructure and technical principles can be measured and potentially applied for verification measures and what degree of explanatory power each specific parameter can provide. It also needs to be considered at which "level"[11] within the IT infrastructure the measure can be performed and to what extent this needs any kind of hardware or software alteration. With regard to the applicability and the political acceptance of possible verification regimes, the following analysis concentrates on parameters and measures that "look from outside" on IT systems and the networks and does not require an alteration of existing IT hardware or software infrastructures. On the other hand, this possibly limits its explanatory power.

[10]It is important to mention, that trade regulation of hardware still can be performed based on the political intent of state actors. But the argumentation for such steps cannot be based on any kind of dual-use considerations.

[11]The term "level" describes the aspect that IT infrastructure and especially networks can be examined at different points and with different amounts of intrusion. As an example, it is non-intrusive to use conventional firewall or monitoring hardware to control the data stream from or to networks at its interconnections with other networks by integrating the hardware into the existing structure. On the other hand, modifying the network structure or even enforcing the usage of specific modified network software will require more extensive adjustments.

The first set of measurable parameters applies to the extent of the hardware of IT systems and networks. Compared to later discussed usage centric monitoring these parameters are quite rough and will not be applicable to monitor the day-to-day usage of the IT systems nor the real time activities of treaty parties like clandestine cyber operations because they represent the overall size of a facility. On the other hand, these parameters are physically obvious, hard to disguise or manipulate and visible for monitoring. They qualify for roughly estimating the storage or processing capacities and to monitor the tendency of the technological developments of facilities as well as reveal the establishment of new cyber capacities or similar significant changes[12]. These parameters are:

- The total power supply as well as the current power consumption of IT infrastructures
- The available supply of cooling systems and their thermal power as well as the current heat production of IT infrastructures
- The available network bandwidth capacities as well as the current flow-rate of transmitted data over monitored network connections
- The total amount of connections of monitored networks to other external civil or commercial networks (the so-called **peering**) and their maximal possible transmission performance
- The amount of required staff for the maintenance of the IT systems.

Beside these lists, other parameters like the CPU and the network processing power as well as the available storage capacities could be used as parameter. But as already pointed out, these are harder to gather because measuring these values needs direct access of monitoring personnel to all controlled systems.

A second set of parameters applies to the usage of IT systems and aims to measure or monitor their specific application. These parameters therefore qualify for the real-time control of cyber operations and activities. In terms of necessary adjustments of the infrastructure, these parameters also can be gathered "from outside" by extending existing infrastructures without the need for any alteration. Nevertheless, in terms of intrusiveness, these parameters are capable of monitoring cyber activities in detail but can contain potentially unwanted or even secret information. These parameters are:

- The metadata of incoming and outbound networked based data transmissions of monitored networks
- The usage of anonymization services
- The usage of exploits for known security problems of IT devices and software

[12] As an example, the analysts of the so called Mandiant report (Mandiant Corporation, 2013) monitored among other parameters the extension of network bandwidth capacities and the necessary infrastructures in Beijing. They used their observations to harden their conclusion, that the Chinese army hosts one of its cyber units, the so called PLA unit 61398, in this area that is suspected to have been the attacker behind many cyber incidents against US companies.

17.4.2 *Approaches for Verification Measures in Cyberspace*

The previous chapter showed that IT systems in fact provide measurable parameters that can be used to develop and establish monitoring procedures. For their deployment, three important aspects need to be considered that affect their technical applicability as well as the potential political acceptance of these measures by treaty parties. These aspects are:

- The technical steps to establish the monitoring systems into existing infrastructures
- The possibly required technical modifications on the monitored systems
- The implementation and maintenance costs.

With regard to a valid estimation of these aspects as well as the practicability of developing monitoring methods it is advisable to analyze existing IT methods from other use cases and possibly adapt them to the new context of verification in contrast to developing measures "on the greenfield". This approach is especially fertile for the cyberspace domain due to the already discussed dual use character of its technologies where the long history of IT security research often already dealt with problems that share similarities to verification problems.

As to the parameters of determining the power supply and cooling capacities of IT infrastructure as well as measuring its actual values this applies to engineering problems that go beyond the scope of this paper and are well understood and established. The same applies to the determination of current and potential network bandwidth capacities and current flow rates, because these things are at the core of safety as well as operating monitoring tasks for data centers. All of these measuring technologies are in the most cases already part of existing IT infrastructure installations, get already logged and don't need any further adjustments except for the aspect of tamperproof storage of the logged data that will be discussed later on. As pointed out, values of these parameters need to get collected and stored over a relevant time because their primary explanatory power lies in the indication of significant infrastructure changes.

A more detailed monitoring of activities needs information about the specific operations that had been and are being performed with IT facilities. This kind of monitoring can be accomplished with methods that acquire and control the usage of specific IT systems or networks. This acquisition is possible on different levels of intrusion. A light-weight version can gather so-called **metadata of outbound and inbound network connections**. This metadata is information which is delivered with the actual payload and always contains at least the IP addresses of the sender and recipient of the transmitted data, the amount of the transmitted data as well as the timestamp of the connection - much like the labels on an envelope. Such types of data already exist because it is necessary for the basic principles of network-based data transmission and processed by all involved networking hard- and software. As for the former discussion, it is therefore merely a question of logging this information, a task which is often already put in place for IT

security or law enforcement reasons[13]. This monitoring of transmitted data could also be intensified if necessary for verification reasons by detecting more in-depth information of the data like the type and content of the data. Such technology is already available and called **deep packet inspection** (Amir, 2007). Gathering and storing such information is always a critical task where personal rights and privacy aspects need to be weighed up against the purpose of this information collection. To respect this, the mentioned storage techniques allow fine-grained possibilities of anonymizing the information to balance the verification agreements on the one hand with the necessities of personal rights, national security and state sovereignty on the other hand. For instance, this would involve the storage of the connection IP addresses on a network level rather than a device specific level.

An important strategy of many cyber operations is their clandestineness and hiding one's tracks – that are, as explained, per default visible – is a key element of such activities. So-called **anonymization services** like Tor, the "onion router network" (Schneier, 1996) provide such services that hide this information, so that connections cannot be attributed to their origin. The principle of such services lies in the routing of any internet connection over specific servers that, in theory, remove any information which would allow to trace it back. Such anonymization networks often utilize a "cloud" of different hubs where connections are additionally routed over to disguise their path. These "disguise clouds" use different cryptographic technologies in a way that the endpoint of the connection does not have any information about its origin. Anonymization technologies undermine effectively the approach of linking cyber operations to their origin and therefore provide a possibility to avoid verification measures. On the other hand, the weak spots of these anonymization services are the entry points, meaning the servers that connect the "disguise cloud" with regular networks. Using the described verification approaches of logging the connections can at least reveal that anonymization services are being used by detecting the connections to the Tor network itself or – in combination with traffic content and traffic pattern detection – by detecting that Tor connections are hidden within the regular data connections stream[14].

One more verification measure that effectively can be monitored is the usage of exploits of known flaws and security holes in software and hardware of IT systems over network connections. The knowledge of such flaws and security problems that often apply to specific versions of software or hardware revisions of technical products are an important source for IT security measures and commonly shared in dedicated databases like the **Common Vulnerabilities and Exposures** (CVE) database. Exploiting these flaws in many cases involves the usage of specific "hand-crafted" network traffic that addresses the security hole at the receiving IT system and triggers purposeful faulty behavior on these IT systems – mostly the bypass of established security measures. These so-called **exploits** can be detected via the traffic analysis methods discussed above when combined with resources like the CVE database (Pimenta Rodrigues et al., 2017). This approach particularly applies to known vulnerabilities and therefore the usage of

[13] An example is provided by the data retention laws in different countries (EU-Parliament, 2006) that are either active per default to store information on internet connections at the servers of IT service providers for a specific time or apply measures that collect this information for the purpose of law enforcement after a court order.

[14] Tor is designed to blend in with regular data traffic and look like normal HTTPS connections. On the other hand, tools that track network traffic and analyze its patterns are able to uncover Tor connections by statistical analysis and due to specific traffic patterns of anonymized connections. An in-depth analysis on this flaw is given by Granerud (Granerud, 2010).

unknown vulnerabilities – so-called **zero-day exploits** –cannot be monitored directly. On the other hand, verification often happens based on stored logged information that is collected over a specific time span and analyzed later. Even while recent studies showed that zero-day exploits often stay undetected for several years[15], this provides at least an approach to put the activities of actors under observation. It must also be seen under the perspective that, as stated before, the most cases of malicious cyber activities do not involve the expensive method of obtaining zero-day vulnerabilities but predominantly exploit existing and well-known security problems (see Verizon, 2018).

17.4.3 *Implementation of Verification Measures*

An important question with regard to the described current state of verification measures for the cyberspace is the question of which existing IT technologies from other use cases can be adopted for this kind of approach. In this case, the dual-use character of the cyberspace can be an advantage, because the necessity of monitoring networks and data connections is also given for IT security reasons and has been a key task since the early days of commercial applications of IT systems. Therefore, a lot of technological developments had been established that can used and its merely a question how the results of these monitoring measures are interpreted. Where IT security aims to detect unwanted intrusions or malicious activities that try to infiltrate a network from the outside, the purpose of verification measures is to detect forbidden activities in terms of the regime agreements, within or from this network. With this in mind, the measuring methods of gathering network connection logs introduced above as well as the more intrusive method of traffic analysis and traffic data inspections and the storage and analysis of this information are "everyday tools" and technologies that are widely used and shall therefore be omitted here. From this point of view, the most critical aspect when it comes to adopting these technologies for verification is the validity of the logged information and its tamper-proof storage.

Ensuring tamper-proof data storage is a problem that can be solved with a relative new technology called **blockchain**. A blockchain "*is a tamper-proof, shared digital ledger that records transactions in a public or private peer-to-peer network. Distributed to all member nodes in the network, the ledger permanently records, in blocks, the history of asset exchanges that take place . . . "* (Iansiti & Lakhani, 2017) and where each block contains a cryptographic hash of the previous block (Purdon & Erturk, 2017). A "hash" can be seen as a technical way of "sealing" information that can be used to ensure for any kind of delivered data that it has not been modified. In the blockchain, each new data entry is verified by its previous entries via a process of so-called **cryptographic signatures**[16]. This means that a digital key is created based on previous entries and then used to cryptographically sign the new entry. This prevents any alteration of stored data because any modification would invalidate all following entries in the blockchain. To ensure that the mechanism that stores the data into the

[15] See the RAND study (Ablon & Bogart, 2017) as an example. The study calculated an average life span of 6.9 years for zero-day exploits. This is put into perspective by other key findings of the study that "*only 25 percent of vulnerabilities do not survive to 1.51 years, and only 25 percent live more than 9.5 years [and that for] a given stockpile of zero-day vulnerabilities, after a year, approximately 5.7 percent have been publicly discovered and disclosed by another entity*".

[16] A brief overview of digital and cryptographic signatures is given (Buchmann, 2004).

blockchain itself is valid and not manipulated, its code or at least a hash of its code can be put into the blockchain for validation. In terms of the defined requirements for the proposed measures, creating and securing logged data with a blockchain mechanism results in a significant increase of the necessary processing and the storage capacities. Nevertheless, using this kind of technical verification for streams of logging data is a concept that had already been described as “audit log” or “audit trail” for use cases in safety or security critical scenarios by Schneier and Kelsey (Schneier & Kelsey, 1998) and is ready to get implemented.

17.5 Conclusion and Outlook

The discussion above has demonstrated the problem of the militarization of the cyberspace and the need for appropriate agreements and accompanying tools of arms control to stabilize this development.

- Verification is one of the pillars for treaties and regimes that enables members or an authorized institution to control each other’s compliance and guarantees the treaties’ effectiveness. While verification as a tool itself has been developed over the last decades for different other technological developments that had been used for military purpose, its application on the cyberspace itself becomes complicated due to the specific features of this new domain. This requires the development of new approaches that, in theory, would result in an ideally tailorable space where mankind can define the rules.
- The previous sections have provided an overview of which existing parameters of this domain are applicable for monitoring and measuring approaches. As demonstrated, such measurement does not require specific technical developments or even specific adjustments of IT infrastructures because they are mostly already installed for IT security reasons.
- This provides an optimistic position for both the establishment of first real-world use cases as well as the further development of such verification measures. For this matter, future work had to been put into the question of how significant the monitoring of specific values is, especially due to the fact that some discussed measurable parameters are mere generic values.
- With regard to the rapid technological development in the field of IT it is also advisable to further analyze how verification measures and their critical thresholds that are monitored can adjust to these developments[17] to reflect its security and stability building intent.
- Further research is also necessary to answer the question of how measures can be developed or strengthened to prevent the circumvention or manipulation of monitoring. And finally, all verification measures are used for specific purposes and use cases. We will soon need to evaluate the proposed measures and find appropriate approaches for the specific tasks, challenges and usage scenarios.

[17]For instance, a simplified and exemplary limit of an energy supply of 10 kilowatts for a facility can generate a multiple of computer processing powers over several years.

LIST OF FIGURES

16.3 Boolean Circuit design output identifier keys that are dependent if their respective vulnerability identifier occurs at least twice. The complexity of the circuit is $\sim 2Nu\ \ 2N^2u\sigma \log(Nu)$ ANDs for number of parties N, number of inputs per party u and security parameter σ. The design is based on Huang *et al.*'s Boolean circuit for computing the set intersection between two parties (Huang et al., 2012). The red values $l_1, \ldots, l_n$ are opened and verified before the oblivious merge block is executed. 245

LIST OF TABLES

BIBLIOGRAPHY

Ablon, L., & Bogart, A. (2017, March 8). *Zero Days, Thousands of Nights: The Life and Times of Zero-Day Vulnerabilities and Their Exploits*. RAND Corporation. https://doi.org/10.7249/RR1751

Aceto, G., Botta, A., Marchetta, P., Persico, V., & Pescapé, A. (2018). A comprehensive survey on internet outages. *Journal of Network and Computer Applications*, *113*, 36–63. https://doi.org/10.1016/j.jnca.2018.03.026

AfTerFibre. (2023). Terestrial fibre optic cables. https://afterfibre.nsrc.org/

Altmann, J. (2019a). Confidence and Security Building Measures for Cyber Forces. In C. Reuter (Ed.), *Information technology for peace and security: IT applications and infrastructures in conflicts, crises, war, and peace* (pp. 185–203). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-25652-4_9

Altmann, J. (2019b). Der Cyber-Rüstungswettlauf. In I.-J. Werkner & N. Schörnig (Eds.), *Cyberwar – die Digitalisierung der Kriegsführung: Fragen zur Gewalt. Band 6* (pp. 87–103). Springer Fachmedien. https://doi.org/10.1007/978-3-658-27713-0_5

Aly, A., Keller, M., Rotaru, D., Scholl, P., Smart, N. P., & Wood, T. (2018). *SCALE-MAMBA*. https://homes.esat.kuleuven.be/~nsmart/SCALE/

Amadae, S. M. (2016). *Prisoners of Reason: Game Theory and Neoliberal Political Economy*. Cambridge University Press. https://doi.org/10.1017/CBO9781107565258

Amarasinghe, A., Wijesinghe, W., Nirmana, D., Jayakody, A., & Priyankara, A. (2019). AI based cyber threats and vulnerability detection, prevention and prediction system. *2019 International Conference on Advancements in Computing (ICAC)*, 363–368. https://doi.org/10.1109/ICAC49085.2019.9103372

Amir, E. (2007). The case for deep packet inspection. *IT Business Edge*. https://www.itbusinessedge.com/

Anati, I., Gueron, S., Johnson, S., & Scarlata, V. (2013). Innovative technology for CPU based attestation and sealing. *Hardware and Architectural Support for Security and Privacy (HASP'13)*, 7. https://www.intel.com/content/dam/develop/external/us/en/documents/hasp-2013-innovative-technology-for-attestation-and-sealing-413939.pdf

Appelbaum, J., Horchert, J., Reißmann, O., Rosenbach, M., Schindler, J., & Stöcker, C. (2013). Neue Dokumente: Der geheime Werkzeugkasten der NSA [newspaper]. *Der Spiegel: Netzwelt*. https://www.spiegel.de/netzwelt/netzpolitik/neue-dokumente-der-geheime-werkzeugkasten-der-nsa-a-941153.html

Apruzzese, G., Colajanni, M., Ferretti, L., & Marchetti, M. (2019). Addressing Adversarial Attacks Against Security Systems Based on Machine Learning. *2019 11th International Conference on Cyber Conflict (CyCon)*, *900*, 1–18. https://doi.org/10.23919/CYCON.2019.8756865

Archer, D. W., Bogdanov, D., Lindell, Y., Kamm, L., Nielsen, K., Pagter, J. I., Smart, N. P., & Wright, R. N. (2018). From Keys to Databases—Real-World Applications of Secure Multi-Party Computation. *The Computer Journal*, *61*(12), 1749–1771. https://doi.org/10.1093/comjnl/bxy090

Arimatsu, L. (2012). A treaty for governing cyber-weapons: Potential benefits and practical limitations. *2012 4th International Conference on Cyber Conflict (CYCON 2012)*, 1–19. https://ieeexplore.ieee.org/document/6243968/authors#authors

Axel, F. (2020, August 28). A Comprehensive Look at Emotet's Summer 2020 Return. https://www.proofpoint.com/us/blog/threat-insight/comprehensive-look-emotets-summer-2020-return

Axelrod, R., & Keohane, R. O. (1985). Achieving Cooperation under Anarchy: Strategies and Institutions. *World Politics*, *38*(1), 226–254. https://doi.org/10.2307/2010357

Baezner, M. (2018, August). *Regional rivalry between India-Pakistan: Tit-for-tat in cyberspace* (Report No. 10). ETH Zurich. https://doi.org/10.3929/ethz-b-000314582

Baezner, M. (2019, May). *Iranian Cyber-activities in the Context of Regional Rivalries and International Tensions* (Report). ETH Zurich. Zurich. https://doi.org/10.3929/ethz-b-000344841

Bahmani, R., Barbosa, M., Brasser, F., Portela, B., Sadeghi, A.-R., Scerri, G., & Warinschi, B. (2017). Secure Multiparty Computation from SGX. In A. Kiayias (Ed.), *Financial Cryptography and Data Security* (pp. 477–497). Springer International Publishing. https://doi.org/10.1007/978-3-319-70972-7_27

Bajema, N. (2019, November 12). *Can Humans Resist the Allure of Machine Speed for Nuclear Weapons?* Outrider. https://outrider.org/nuclear-weapons/articles/can-humans-resist-allure-machine-speed-nuclear-weapons

Ballweber, J., & Reinhold, T. (2022). Starlink in ukraine: The cosmic whims of a billionaire [newspaper]. *Frankfurter Rundschau: Politik*. https://www.fr.de/politik/ukraine-krieg-starlink-elon-musk-kosmische-launen-milliardaer-netzwerk-news-91863875.html.

Baram, G., & Sommer, U. (2019). Covert or not Covert: National Strategies During Cyber Conflict. *2019 11th International Conference on Cyber Conflict (CyCon)*, *900*, 1–16. https://doi.org/10.23919/CYCON.2019.8756682

Bargar, A., Pitts, S., Butkevics, J., & McCulloh, I. (2019). Challenges and opportunities to counter information operations through social network analysis and theory. *2019 11th International Conference on Cyber Conflict (CyCon*, *900*, 1–18. https://doi.org/10.23919/CYCON.2019.8756832

Barrat, A., Barthélemy, M., & Vespignani, A. (2007, June). The Architecture of Complex Weighted Networks: Measurements and Models. In *Large Scale Structure and Dynamics of Complex Networks* (pp. 67–92, Vol. Volume 2). WORLD SCIENTIFIC. https://doi.org/10.1142/9789812771681_0005

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Batemann, J. (2022). Russia's wartime cyber operations in ukraine: Military impacts, influences, and implications. *Carnegie Endowment for International Peace December*, *16*. https://carnegieendowment.org/2022/12/16/russia-s-wartime-cyber-operations-in-ukraine-military-impacts-influences-and-implications-pub-88657.

Bazin, A. (2013, July 31). *Winning Trust and Confidence: A Grounded Theory Model for the Use of Confidence-Building Measures in the Joint Operational Environment*.

https://www.academia.edu/8104468/Winning_Trust_and_Confidence_A_Grounded_Theory_Model_for_the_Use_of_Confidence_Building_Measures_in_the_Joint_Operational_Environment

Beaver, D., Micali, S., & Rogaway, P. (1990). The round complexity of secure protocols (extended abstract). *Symposium on Theory of Computing (STOC'90)*, 503–513. https://www.cs.ucdavis.edu/~rogaway/papers/bmr90

Becker, U., Müller, H., & Rosert, E. (2008). Einleitung: Rüstungskontrolle im 21. Jahrhundert. *Die Friedens-Warte*, *83*(2/3), 13–34. https://www.jstor.org/stable/23773893

Beecroft, N. (2022). Evaluating the international support to Ukrainian cyber defense. *Carnegie Endowment for International Peace November*, *3*. https://carnegieendowment.org/2022/11/03/evaluating-international-support-to-ukrainian-cyber-defense-pub-88322.

Bellare, M., & Rogaway, P. (2005). Introduction to modern cryptography. http://web.cs.ucdavis.edu/~rogaway/classes/227/spring05/book/main.pdf

Bendiek, A. (2016). Sorgfaltsverantwortung im Cyberraum - Leitlinien für eine deutsche Cyber-Außen- und Sicherheitspolitik. https://www.swp-berlin.org/publications/products/studien/2016S03_bdk.pdf

Bendiek, A., & Schulze, M. (2021). *Attribution als Herausforderung für EU Cybersanktionen*. SWP. https://doi.org/10.1007/978-3-658-25652-4_10

Bendlin, R., Damgård, I., Orlandi, C., & Zakarias, S. (2011). Semi-homomorphic Encryption and Multiparty Computation. In K. G. Paterson (Ed.), *Advances in Cryptology – EUROCRYPT 2011* (pp. 169–188). Springer. https://doi.org/10.1007/978-3-642-20465-4_11

Biehl, H., & Jacobs, J. (2014). Öffentliche Meinung und Sicherheitspolitik. In S. Böckenförde & S. Gareis (Eds.), *Deutsche Sicherheitspolitik. Herausforderungen, Akteure und Prozesse* (2nd ed., pp. 265–274). UTB Verlag. https://www.utb.de/doi/book/10.36198/9783838585116

Biermann, K. (2015). Vorratsdaten kosten mindestens 260 Millionen Euro [newspaper]. *Die Zeit*. https://www.zeit.de/digital/datenschutz/2015-10/vds-vorratsdatenspeicherung-millionen-kosten

Bigelow, B. (2019). What are Military Cyberspace Operations Other Than War? *2019 11th International Conference on Cyber Conflict (CyCon)*, *900*, 1–17. https://doi.org/10.23919/CYCON.2019.8756835

Biller, J., & Schmitt, M. N. (2019, June 19). *Classification of Cyber Capabilities and Operations as Weapons, Means, or Methods of Warfare*. https://papers.ssrn.com/abstract=3424500

Biryukov, A., Dinu, D., & Khovratovich, D. (2016). Argon2: New generation of memory-hard functions for password hashing and other applications. *European Symposium on Security and Privacy (EuroS&P)*, 292–302. https://doi.org/10.1109/EuroSP.2016.31

Bischof, Z. S., Fontugne, R., & Bustamante, F. E. (2018). Untangling the world-wide mesh of undersea cables. *Proceedings of the 17th ACM Workshop on Hot Topics in Networks*, 78–84. https://doi.org/10.1145/3286062.3286074

Blagden, D. (2020). Deterring Cyber Coercion: The Exaggerated Problem of Attribution. *Survival*, *62*(1), 131–148. https://doi.org/10.1080/00396338.2020.1715072

Boehme, P. (2008). The Verification Regime of the Chemical Weapons Convention: An Overview. https://www.opcw.org/media-centre/news/2008/11/verification-regime-chemical-weapons-convention-overview

Bogdanov, D., Laur, S., & Willemson, J. (2008). Sharemind: A Framework for Fast Privacy-Preserving Computations. In S. Jajodia & J. Lopez (Eds.), *Computer Security - ESORICS 2008* (pp. 192–206). Springer. https://doi.org/10.1007/978-3-540-88313-5_13

Bogetoft, P., Christensen, D. L., Damgård, I., Geisler, M., Jakobsen, T., Krøigaard, M., Nielsen, J. D., Nielsen, J. B., Nielsen, K., Pagter, J., Schwartzbach, M., & Toft, T. (2009). Secure Multiparty Computation Goes Live. In R. Dingledine & P. Golle (Eds.), *Financial Cryptography and Data Security* (pp. 325–343). Springer. https://doi.org/10.1007/978-3-642-03549-4_20

Bosch Global. (2021). *Carbyne stack: Cloud native secure multiparty computation*. https://www.bosch.com/stories/open-source-carbyne-stack/

Boulanin, V. (2019, May). *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk, Volume I, Euro-Atlantic perspectives*. SIPRI. https://www.sipri.org/publications/2019/other-publications/impact-artificial-intelligence-strategic-stability-and-nuclear-risk-volume-i-euro-atlantic

Bradner, S. (1999). Internet engineering task force. In *Open sources: Voices from the open source revolution* (pp. 47–52). O'Reilly & Associates. https://www.oreilly.com/openbook/opensources/book/ietf.html

Braun, L., Demmler, D., Schneider, T., & Tkachenko, O. (2022). MOTION – A Framework for Mixed-Protocol Multi-Party Computation. *ACM Transactions on Privacy and Security*, *25*(2), 8:1–8:35. https://doi.org/10.1145/3490390

Bright, A. (2007). Estonia accuses Russia of 'cyberattack' [magazine]. *Christian Science Monitor*. https://www.csmonitor.com/2007/0517/p99s01-duts.html

Brockmann, K. (2019). Challenges To Multilateral export Controls. The Case for Inter-regime Dialogue and Coordination. https://www.sipri.org/publications/2019/other-publications/challenges-multilateral-export-controls-case-inter-regime-dialogue-and-coordination

Brodkin, J. (2015, September 7). *Broken cable reportedly disconnected US island territory from Internet*. Ars Technica. https://arstechnica.com/information-technology/2015/07/broken-cable-reportedly-disconnected-us-island-territory-from-internet/

Broeders, D., Busser, E. D., & Pawlak, P. (2020). Three tales of attribution in cyberspace. https://eucyberdirect.eu/content_research/three-tales-of-attribution-in-cyberspace/

Broeders, D., & Cristiano, F. (2020, April 2). *Cyber Norms and the United Nations: Between Strategic Ambiguity and Rules of the Road*. https://doi.org/10.2139/ssrn.3819171

Bromium. (2019). *EMOTET: A Technical Analysis of The Destructive Polymorphic Malware*. Bromium. https://www.bromium.com/wp-content/uploads/2019/07/Bromium-Emotet-Technical-Analysis-Report.pdf

Bronk, C., & Tikk-Ringas, E. (2013). The Cyber Attack on Saudi Aramco. *Survival*, *55*(2), 81–96. https://doi.org/10.1080/00396338.2013.784468

Brown, G. D., & Tullos, O. W. (2012). On the Spectrum of Cyberspace Operations. *Small Wars Journal*. http://smallwarsjournal.com/jrnl/art/on-the-spectrum-of-cyberspace-operations

Brown, J. M., & Fazal, T. M. (2021). #SorryNotSorry: Why states neither confirm nor deny responsibility for cyber operations. *European Journal of International Security*, *6*(4), 401–417. https://doi.org/10.1017/eis.2021.18

Brühl, T. (2019). Krise des Multilateralismus – Krise der Vereinten Nationen? https://zeitschrift-vereinte-nationen.de/suche/zvn/artikel/krise-des-multilateralismus-krise-der-vereinten-nationen

Buchan, R. (2018). *Cyber Espionage and International Law*. Hart Publishing. https://doi.org/10.1017/S0922156519000359

Buchanan, B., & Cunningham, F. S. (2020). Preparing the Cyber Battlefield: Assessing a Novel Escalation Risk in a Sino-American Crisis (Fall 2020). *Texas National Security Review*, *3*(4), 54–81. https://doi.org/10.26153/tsw/10951

Buchmann, J. A. (2004). Digital signatures. In *Introduction to cryptography* (pp. 249–275). Springer New York. https://doi.org/10.1007/978-1-4419-9003-7_12

Bueger, C., & Liebetrau, T. (2021). Protecting hidden infrastructure: The security politics of the global submarine data cable network. *Contemporary Security Policy*, *42*(3), 391–413. https://doi.org/10.1080/13523260.2021.1907129

Bundesnetzagentur. (2015). Gesetzentwurf zur Speicherpflicht - Kostenschätzung. https://gruen-digital.de/wp-content/uploads/2015/10/BNetzA.pdf

Burgers, T., & Robinson, D. R. S. (2018). Keep Dreaming: Cyber Arms Control is Not a Viable Policy Option. *S&F Sicherheit und Frieden*, *36*(3), 140–145. https://doi.org/10.5771/0175-274X-2018-3-140

Burton, J., & Christou, G. (2021). Bridging the gap between cyberwar and cyberpeace. *International Affairs*, *97*(6), 1727–1747. https://doi.org/10.1093/ia/iiab172

Burton, J., & Soare, S. R. (2019). Understanding the Strategic Implications of the Weaponization of Artificial Intelligence. *2019 11th International Conference on Cyber Conflict (CyCon)*, *900*, 1–17. https://doi.org/10.23919/CYCON.2019.8756866

Butts, C. (2019). *Tools for social network analysis*. https://cran.r-project.org/package=sna

Caine, K. (2016). Local Standards for Sample Size at CHI. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 981–992. https://doi.org/10.1145/2858036.2858498

Calle, E., Cosgaya, S. G., Martínez, D., & Pióro, M. (2019). Solving The Backup Controller Placement Problem In SDN Under Simultaneous Targeted Attacks. *2019 11th International Workshop on Resilient Networks Design and Modeling (RNDM)*, 1–7. https://doi.org/10.1109/RNDM48015.2019.8949096

Campbell, K. M., Einhorn, R. J., Reiss, M. B., & Gregorian, V. (2005). Avoiding the Tipping Point: Concluding Observations. In *The Nuclear Tipping Point: Why States Reconsider Their Nuclear Choices* (pp. 317–348). Brookings Institution Press. https://muse.jhu.edu/pub/11/edited_volume/book/52029

Cannellos, M., & Haga, R. (2016). Lost in Translation: Getting Autonomous Weapons Systems Ethicists, Regulators, and Technologists to Speak the Same Language. *IEEE Technology and Society Magazine*, *35*(3), 50–58. https://doi.org/10.1109/MTS.2016.2593218

Cariolle, J. (2018). *Telecommunication Submarine-Cable Deployment and the Digital Divide in Sub-Saharan Africa*. https://doi.org/10.2139/ssrn.3338769

Carlini, N., & Wagner, D. (2014). {ROP} is Still Dangerous: Breaking Modern Defenses, 385–399. https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/carlini

Carter, L., Burnett, D., Drew, S., Hagadorn, L., Marle, G., Bartlett-Mcneil, D., & Irvine, N. (2009). Submarine cables and the oceans: Connecting the world. *UNEP-WCMC Biodiversity Ser.*, *31*. https://www.researchgate.net/publication/286143047_Submarine_cables_and_the_oceans_Connecting_the_world

Carvalho, M., & Ford, R. (2014). Moving-Target Defenses for Computer Networks. *IEEE Security & Privacy*, *12*(2), 73–76. https://doi.org/10.1109/MSP.2014.30

CCDCOE. (2013). *Tallinn Manual on the International Law Applicable to Cyber Warfare* (M. N. Schmitt, Ed.). Cambridge University Press. https://www.cambridge.org/core/books/tallinn-manual-on-the-international-law-applicable-to-cyber-warfare/50C5BFF166A7FED75B4EA643AC677DAE

CCDCOE. (2017). *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (M. Schmitt & L. Vihul, Eds.; 2nd ed.). Cambridge University Press. https://doi.org/10.1017/9781316822524

CCDCOE. (2018). Mitigating risks arising from false-flag and no-flag cyber attacks. https://ccdcoe.org/uploads/2018/10/False-flag-and-no-flag-20052015.pdf

CCDCOE. (2020). Recent Cyber Events and Possible Implications for Armed Forces, 1–6. https://ccdcoe.org/uploads/2020/09/Recent-Cyber-Events-and-Possible-Implications-for-Armed-Forces-5-September-2020_Final.pdf

CCDCOE. (2022). Cyberspace strategic outlook 2030. NATO cooperative cyber defence centre of excellence. https://ccdcoe.org/library/publications/cyberspace-strategic-outlook-2030-horizon-scanning-and-analysis/

CCM. (2008). The Convention on Cluster Munitions (CCM). https://www.clusterconvention.org/convention-text/

CFE. (1990). Treaty on conventional armed forces in europe. https://www.osce.org/library/14087?download=true

Charniak, E. (2019). *Introduction to Deep Learning*. The MIT Press. https://dl.acm.org/doi/book/10.5555/3351847

Chaudhari, H., Choudhury, A., Patra, A., & Suresh, A. (2019). ASTRA: High Throughput 3PC over Rings with Application to Secure Prediction. *Proceedings of the 2019 ACM SIGSAC Conference on Cloud Computing Security Workshop*, 81–92. https://doi.org/10.1145/3338466.3358922

Chaudhari, H., Rachuri, R., & Suresh, A. (2020). Trident: Efficient 4PC Framework for Privacy Preserving Machine Learning. *Proceedings 2020 Network and Distributed System Security Symposium*. https://doi.org/10.14722/ndss.2020.23005

Cheikes, B. A., Kent, K. A., & Waltermire, D. (2011). *Common platform enumeration: Naming specification version 2.3*. US Department of Commerce, National Institute of Standards and Technology. https://nvlpubs.nist.gov/nistpubs/Legacy/IR/nistir7695.pdf?ref=https://githubhelp.com

Chen, S., Sun, S., & Kang, S. (2020). System integration of terrestrial mobile communication and satellite communication —the trends, challenges and key technologies in B5G and 6G. *China Communications*, *17*(12), 156–171. https://doi.org/10.23919/JCC.2020.12.011

Chen, W., & Wellman, B. (2004). The global digital divide–within and between countries. *IT & society*, *1*(7), 18–25. https://d1wqtxts1xzle7.cloudfront.net/30596463/10-CHEN_AND_WELLMAN_-_GLOBAL_DIGITAL_DIVIDE-libre.pdf?1391847227=&response-content-disposition=inline%3B+filename%3D2004_The_Global_Digital_Divide_Within_A.pdf&Expires=1681233714&Signature=Ld81adj2hTbygTiQCdvZsWQs9BLnDt92HmRbBSrExGy6WEAeZ~yp5EU~n2fnG5eoQqBRXFmC1dm92DCxTEA93fyI2luK18he36MX5Jj0IcigMCnNyJzblBOztExLBRfDHL0YutrFpencdkWFFdVGIIwq2S~76TfOjU91M8OosCWrv-8hXSJ0qyb00M9lQqiUh676VVYoUE7aiwimtq77iXtUHnko9tpyru570nX8X5OIJjTPJnniBgX~QBLnmiKkTCUGKCHcPeVLoHbqrGE8gp3T68

BPDc-FXPEJROdrAO5NzgCeeeYil-gkvHw9jgUGGiYGVeQyBk6xwbnSDVko4w__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA
Chirgwin, R. (2018). IT 'Heroes' saved maersk from NotPetya with ten-day reinstallation blitz. https://www.theregister.co.uk/2018/01/25/after_notpetya_maersk_replaced_everything/.
Chohan, U. W. (2019, February 20). *The Limits to Blockchain? Scaling vs. Decentralization*. https://doi.org/10.2139/ssrn.3338560
CIA. (2021). The world factbook. https://www.cia.gov/the-world-factbook/
Cimpanu, C. (2019). *Kaspersky identifies mysterious APT mentioned in 2017 Shadow Brokers leak*. ZDNET. https://www.zdnet.com/article/kaspersky-identifies-mysterious-apt-mentioned-in-2017-shadow-brokers-leak/
CISA. (2020). *Emotet Malware - Alert (TA18-201A)*. CISA Reports. https://us-cert.cisa.gov/ncas/alerts/TA18-201A
Clark, D. D., & Landau, S. (2010). The problem isn't attribution: It's multi-stage attacks. *Proceedings of the Re-Architecting the Internet Workshop*, 1–6. https://doi.org/10.1145/1921233.1921247
Clark, D. D., & Landau, S. (2011). Untangling Attribution. *Harvard National Security Journal*, *2*, 323. https://heinonline.org/HOL/Page?handle=hein.journals/harvardnsj2&id=531&div=&collection=
Cohen, F. (1987). Computer viruses: Theory and experiments. *Computers & Security*, *6*(1), 22–35. https://doi.org/10.1016/0167-4048(87)90122-2
Commonwealth Secretariat. (2018). Commonwealth Cyber Declaration. https://thecommonwealth.org/commonwealth-cyber-declaration-2018
Corera, G. (2022). Inside a US military cyber team's defence of Ukraine [newspaper]. *BBC News: World*. https://www.bbc.com/news/uk-63328398
Coughlin, T. (2020). HDD market history and projections. *Forbes*. https://www.forbes.com/sites/tomcoughlin/2020/05/29/hdd-market-history-and-projections/?sh=7f2ddb986682
Council on Foreign Relations. (2023). *Cyber Operations Tracker*. Cyber Operations Tracker. https://www.cfr.org/cyber-operations/
Cox, J. (2022, March 28). *Video: Belarusian Cyber Partisans Explain Why They're Hacking to Stop Russia*. Vice. https://www.vice.com/en/article/m7vwxq/video-belarusian-cyber-partisans-explain-why-theyre-hacking-to-stop-russia
Croman, K., Decker, C., Eyal, I., Gencer, A. E., Juels, A., Kosba, A., Miller, A., Saxena, P., Shi, E., Sirer, E. G., Song, D., & Wattenhofer, R. (2016). On scaling decentralized blockchains. *Proceedings of the International Conference on Financial Cryptography and Data Security FC 2016*, *9604 LNCS*, 106–125. https://doi.org/10.1007/978-3-662-53357-4_8
CrowdStrike. (2016). *Who Is COZY BEAR?* https://www.crowdstrike.com/blog/who-is-cozy-bear/
Crysys Inc. (2012). sKyWIper (a.k.a. Flame a.k.a. Flamer): A complex malware for targeted attacks. https://www.crysys.hu/publications/files/skywiper.pdf
Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. http://igraph.org
CSCE. (1975). Conference on security and cooperation in Europe: Final Act. https://www.osce.org/helsinki-final-act
CWC. (1997). Convention on the prohibition of the development, production, stockpiling and use of chemical weapons and on their destruction. https://www.opcw.org/chemical-weapons-convention

Czosseck, C., & Podins, K. (2012). A Vulnerability-Based Model of Cyber Weapons and its Implications for Cyber Conflict. *International Journal of Cyber Warfare and Terrorism (IJCWT)*, *2*(1), 14–26. https://doi.org/10.4018/ijcwt.2012010102

Daase, C., Neuneck, G., Schaper, A., Schmidt, H.-J., & Wunderlich, C. (2019). Nukleare Weltordnung in der Krise. In H. BICC & I. IFSH (Eds.), *2019. Vorwärts in die Vergangenheit? Frieden braucht Partner. Friedensgutachten* (pp. 25–43). LIT Verlag. https://friedensgutachten.de/user/pages/04.archiv/2019/02.ausgabe/03.fokus/FGA_2019_Fokus.pdf

Damgård, I., Keller, M., Larraia, E., Pastro, V., Scholl, P., & Smart, N. P. (2013). Practical Covertly Secure MPC for Dishonest Majority – Or: Breaking the SPDZ Limits. In J. Crampton, S. Jajodia, & K. Mayes (Eds.), *Computer Security – ESORICS 2013* (pp. 1–18). Springer. https://doi.org/10.1007/978-3-642-40203-6_1

Damgård, I., & Orlandi, C. (2010). Multiparty Computation for Dishonest Majority: From Passive to Active Security at Low Cost. In T. Rabin (Ed.), *Advances in Cryptology – CRYPTO 2010* (pp. 558–576). Springer. https://doi.org/10.1007/978-3-642-14623-7_30

Damgård, I., Pastro, V., Smart, N., & Zakarias, S. (2012). Multiparty Computation from Somewhat Homomorphic Encryption. In R. Safavi-Naini & R. Canetti (Eds.), *Advances in Cryptology – CRYPTO 2012* (pp. 643–662). Springer. https://doi.org/10.1007/978-3-642-32009-5_38

Danchev, D. (2008, August 11). Coordinated russia vs georgia cyberattack in progress. https://www.zdnet.com/article/coordinated-russia-vs-georgia-cyber-attack-in-progress/

Davies, A. (2022, April 19). *Private Blockchain: Implementation Guide I DevTeam.Space*. DevTeam.Space. https://www.devteam.space/blog/private-blockchain-implementation-guide/

Davis II, J. S., Boudreaux, B., Welburn, J. W., Aguirre, J., Ogletree, C., McGovern, G., & Chase, M. S. (2017, June 2). *Stateless Attribution: Toward International Accountability in Cyberspace* (RAND.). RAND Corporation. https://www.rand.org/pubs/research_reports/RR2081.html

Dekker, D., & Karsberg, C. (2014). Technical guideline on incident reporting technical guidance on the incident reporting in article 13a. https://resilience.enisa.europa.eu/article-13/guideline-for-incident-reporting/Article_13a_ENISA_Technical_Guideline_On_Incident_Reporting_v2_1.pdf

Demir, L., Kumar, A., Cunche, M., & Lauradoux, C. (2018). The Pitfalls of Hashing for Privacy. *IEEE Communications Surveys & Tutorials*, *20*(1), 551–565. https://doi.org/10.1109/COMST.2017.2747598

Demmler, D., Schneider, T., & Zohner, M. (2015). ABY - A Framework for Efficient Mixed-Protocol Secure Two-Party Computation. *Proceedings 2015 Network and Distributed System Security Symposium*. https://doi.org/10.14722/ndss.2015.23113

Den Dekker, G. (2004). The Effectiveness of International Supervision in Arms Control Law. *Journal of Conflict and Security Law*, *9*(3), 315–330. https://doi.org/10.1093/jcsl/9.3.315

DeNardis, L. (2012). Hidden Levers of Internet Control. *Information, Communication & Society*, *15*(5), 720–738. https://doi.org/10.1080/1369118X.2012.659199

Denning, D. (2001). Obstacles and options for cyber arms control. *Arms Control in Cyberspace, Heinrich Böll Foundation, Berlin, Germany*, 1–13. http://faculty.nps.edu/dedennin/publications/berlin.pdf

Desouza, K. C., Ahmad, A., Naseer, H., & Sharma, M. (2020). Weaponizing information systems for political disruption: The Actor, Lever, Effects, and Response Taxonomy (ALERT). *Computers & Security*, *88*, 101606. https://doi.org/10.1016/j.cose.2019.101606

Dewar, R. (2017, November 14). *Cyberweapons: Capability, Intent and Context in Cyberdefense* (Report). ETH Zurich. https://doi.org/10.3929/ethz-b-000210449

Dhir, N., Hoeltgebaum, H., Adams, N., Briers, M., Burke, A., & Jones, P. (2021, April 20). *Prospective Artificial Intelligence Approaches for Active Cyber Defence*. https://doi.org/10.48550/arXiv.2104.09981

Dickey, L., Downs, E., Taffer, A., Holz, H., Thompson, D., Hyder, S. B., Loomis, R., & Miller, A. (2019). *Mapping the Information Environment in the Pacific Island Countries: Disruptors, Deficits, and Decisions*. Center for Naval Analysis Arlington United States. https://apps.dtic.mil/sti/citations/AD1115856

Ding, Y., Wu, R., & Zhang, X. (2019). Ontology-based knowledge representation for malware individuals and families. *Computers & Security*, *87*, 101574. https://doi.org/10.1016/J.COSE.2019.101574

Dishington, C., Sharma, D. P., Kim, D. S., Cho, J.-H., Moore, T. J., & Nelson, F. F. (2019). Security and Performance Assessment of IP Multiplexing Moving Target Defence in Software Defined Networks. *2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, 288–295. https://doi.org/10.1109/TrustCom/BigDataSE.2019.00046

Dittrich, P.-J., & Boening, D. (2017). Mehr Sicherheit im Cyberraum. Plädoyer für eine Rüstungskontrolle. *Arbeitspapier Sicherheitspolitik*, *9*, 1–5. https://www.baks.bund.de/sites/baks010/files/baks_arbeitspapier_sicherheitspolitik_2017_09.pdf

Division, U. N. S. (2018). Methodology: Standard country or area codes for statistical use (M49). https://unstats.un.org/unsd/methodology/m49/%20(accessed%20January%2019,%202022)

Doherty, B., & McClure, T. (2022). Tonga could be cut off for weeks amid efforts to repair undersea communications cable [newspaper]. *The Guardian: World news*. https://www.theguardian.com/world/2022/jan/18/tonga-could-be-cut-off-for-weeks-amid-efforts-to-repair-undersea-communications-cable

Dong, Y., Guo, W., Chen, Y., Xing, X., Zhang, Y., & Wang, G. (2019). Towards the Detection of Inconsistencies in Public Security Vulnerability Reports, 869–885. https://www.usenix.org/conference/usenixsecurity19/presentation/dong

Dragos Inc. (2017). TRISIS Malware, 1–19. https://www.energy.senate.gov/public/index.cfm/files/serve?File_id=40B2ED59-D34E-47C3-B9E2-1E8D030C5748

Droz, S., & Stauffacher, D. (2018). *Trust and Attribution in Cyberspace: A proposal for an independent network of organisations engaging in attribution peer-review*. ICT4Peace Foundation. Geneva. https://ict4peace.org/wp-content/uploads/2018/12/ICT4Peace-2019-Trust-and-Attribution-in-Cyberspace.pdf

Duckett, C. (2022). Volcanic eruption takes out Tonga cables. https://www.zdnet.com/article/volcanic-eruption-takes-out-tonga-cables/

Duncan, B. (2021, January 19). *Wireshark Tutorial: Examining Emotet Infection Traffic*. Unit 42. https://unit42.paloaltonetworks.com/wireshark-tutorial-emotet-infection/

Dunn, M. (2005). *A comparative analysis of cybersecurity initiatives worldwide*. Center for Security Studies. https://www.key4biz.it/files/000050/00005073.pdf

Dwyer, A., & Silomon, J. (2019, September 23). *Dangerous Gaming: Cyber-Attacks, Air-Strikes and Twitter*. E-International Relations. https://www.e-ir.info/2019/09/23/dangerous-gaming-cyber-attacks-air-strikes-and-twitter/

Economist. (2010). Cyberwar [magazine]. *The Economist*. https://www.economist.com/leaders/2010/07/01/cyberwar

Eggers, S. (2021). A novel approach for analyzing the nuclear supply chain cyber-attack surface. *Nuclear Engineering and Technology*, *53*(3), 879–887. https://doi.org/10.1016/j.net.2020.08.021

Egloff, F., & Wenger, A. (2019). Public attribution of cyber incidents. *CSS Analyses in Security Policy*, *244*, 4. https://css.ethz.ch/content/dam/ethz/special-interest/gess/cis/center-for-securities-studies/pdfs/CSSAnalyse244-EN.pdf

E.G.N.Y.T.E. (2021). Data retention 101: Policies and best practices. *EGNYTE Blog*. https://www.egnyte.com/guides/governance/data-retention

Ehrenfeld, J. M. (2017). WannaCry, Cybersecurity and Health Information Technology: A Time to Act. *Journal of Medical Systems*, *41*(7), 104. https://doi.org/10.1007/s10916-017-0752-1

Ehrhart, H.-G. (2019). Russia's "hybrid warfare". *W&F, Science and Peace*, *37*(3), 17–19. https://wissenschaft-und-frieden.de/artikel/russlands-hybride-kriegfuehrung/

Eilstrup-Sangiovanni, M. (2018). Why the World Needs an International Cyberwar Convention. *Philosophy & Technology*, *31*(3), 379–407. https://doi.org/10.1007/s13347-017-0271-5

Ekran Systems. (2022). *15 cybersecurity best practices to prevent cyber attacks*. Ekran Systems. https://www.ekransystem.com/en/blog/best-cyber-security-practices

ENISA. (2011). *Cyber Europe 2010 – Evaluation Report* (Evaluation report, 1–47.). https://doi.org/10.2824/218244

ENISA. (2012). Cyber Europe 2012 - Key Findings Report. https://www.enisa.europa.eu/topics/publications/cyber-europe-2012-key-findings-report

ENISA. (2017). ENISA threat landscape report 2017. *Report/Study*. https://www.enisa.europa.eu/publications/enisa-threat-landscape-report-2017

ENISA. (2018). *Reference Incident Classification Taxonomy*. https://www.enisa.europa.eu/publications/reference-incident-classification-taxonomy

ESET. (2018). *One year later: EternalBlue exploit more popular now than during WannaCryptor outbreak*. ESET. https://www.welivesecurity.com/2018/05/10/one-year-later-eternalblue-exploit-wannacryptor/

EU-Commission. (2016a). Commission proposes to modernise and strengthen controls on exports of dual-use items. https://ec.europa.eu/commission/presscorner/detail/en/IP_16_3190

EU-Commission. (2016b). Regulation of the European Parliament and of the Council setting up a Union regime for the control of exports, transfer, brokering, technical assistance and transit of dual-use items (recast). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52016PC0616

EU-Commission. (2022). *Cyber Defence: EU boosts action against cyber threats*. https://ec.europa.eu/commission/presscorner/detail/en/ip_22_6642

EU-Council. (2001). Convention on Cybercrime. *33*(0), 6–8. https://rm.coe.int/1680081561

EU-FRA. (2017). Data retention across the EU. *European Agency for Fundamental Rights*. https://fra.europa.eu/en/publication/2017/data-retention-across-eu#publication-tab-0

EU-Parliament. (2006, March 15). Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC. http://data.europa.eu/eli/dir/2006/24/oj/eng

EU-Parliament. (2008). Directive 2008/114/EC on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32008L0114&from=EN5

Europol, UNICRI, & Trend Micro. (2020). Malicious Uses and Abuses of Artificial Intelligence, 80. https://www.europol.europa.eu/publications-events/publications/malicious-uses-and-abuses-of-artificial-intelligence

Evans, C. (2014, July 15). *Project Zero: Announcing Project Zero*. Project Zero. https://googleprojectzero.blogspot.com/2014/07/announcing-project-zero.html

Evans, D., Kolesnikov, V., & Rosulek, M. (2018). A Pragmatic Introduction to Secure Multi-Party Computation. *Foundations and Trends® in Privacy and Security*, *2*(2-3), 70–246. https://doi.org/10.1561/3300000019

Facebook. (2018). Facebook White-Hat. https://www.facebook.com/whitehat/

Faesen, L., Klimburg, A., Van Hoeve, S., Minicozzi, R., Siemens, S. P., & Tesauro, G. (2022). *The Cyber Arms Watch Uncovering the Stated & Perceived Offensive Cyber Capabilities of States*. https://hcss.nl/wp-content/uploads/2022/05/Cyber-Arms-Watch-HCSS-2022-V.2.pdf

Falliere, N., Murchu, L. O., & Chien, E. (2011). *W32.Stuxnet Dossier V1.4*. Symantec. https://www.wired.com/images_blogs/threatlevel/2011/02/Symantec-Stuxnet-Update-Feb-2011.pdf

Faramondi, L., Oliva, G., & Setola, R. (2020). Multi-criteria node criticality assessment framework for critical infrastructure networks. *International Journal of Critical Infrastructure Protection*, *28*, 100338. https://doi.org/10.1016/j.ijcip.2020.100338

Fayi, S. Y. A. (2018). What Petya/NotPetya Ransomware Is and What Its Remidiations Are. In S. Latifi (Ed.), *Information Technology - New Generations* (pp. 93–100). Springer International Publishing. https://doi.org/10.1007/978-3-319-77028-4_15

Feldmann, A., Gasser, O., Lichtblau, F., Pujol, E., Poese, I., Dietzel, C., Wagner, D., Wichtlhuber, M., Tapiador, J., Vallina-Rodriguez, N., Hohlfeld, O., & Smaragdakis, G. (2020). The Lockdown Effect: Implications of the COVID-19 Pandemic on Internet Traffic. *Proceedings of the ACM Internet Measurement Conference*, 1–18. https://doi.org/10.1145/3419394.3423658

Felsen, S., Kiss, Á., Schneider, T., & Weinert, C. (2019). Secure and Private Function Evaluation with Intel SGX. *Proceedings of the 2019 ACM SIGSAC Conference on Cloud Computing Security Workshop*, 165–181. https://doi.org/10.1145/3338466.3358919

Felton, D. (2019, July 5). *What is a trusted execution environment (TEE)?* Trustonic blog. https://www.trustonic.com/technical-articles/what-is-a-trusted-execution-environment-tee/

Field, M. (2019, December 20). *As the US, China, and Russia build new nuclear weapons systems, how will AI be built in?* Bulletin of the Atomic Scientists.

https://thebulletin.org/2019/12/as-the-us-china-and-russia-build-new-nuclear-weapons-systems-how-will-ai-be-built-in/
FIfF e.V. (2017). *Cyberpeace statt cyberwar!* https://blog.fiff.de/cyberpeace/
Finifter, M., Akhawe, D., & Wagner, D. (2013). An Empirical Study of Vulnerability Rewards Programs, 273–288. https://www.usenix.org/conference/usenixsecurity13/technical-sessions/presentation/finifter
Finnemore, M. (2017). Cybersecurity and the concept of norms. https://carnegieendowment.org/2017/11/30/cybersecurity-and-concept-of-norms-pub-74870
FireEye. (2014). *APT28: A Window Into Russia's Cyber Espionage Operations?* https://www2.fireeye.com/rs/fireye/images/rpt-apt28.pdf
FortiGuard. (2023). *Global Threat Landscape Report A Semiannual Report by FortiGuard Labs*. https://www.fortinet.com/blog/threat-research/fortiguard-labs-threat-report-key-findings-2h-2022
Fouguet, H. (2021). China's 7,500-Mile Undersea Cable to Europe Fuels Internet Feud [newspaper]. *Bloomberg.com*. https://www.bloomberg.com/news/articles/2021-03-05/china-s-peace-cable-in-europe-raises-tensions-with-the-u-s
Franken, J., Reinhold, T., Reichert, L., & Reuter, C. (2022). The Digital Divide in State Vulnerability to Submarine Communications Cable Failure. *International Journal of Critical Infrastructure Protection*, *38*. https://doi.org/10.1016/j.ijcip.2022.100522
Frei, S. (2013). The known unknowns. Empirical analysis of publicly unknown security vulnerabilities. In *NSS labs briefs* (p. 6). NSS Labs. https://techzoom.net/whitepapers/the_known_unknowns_2013.pdf
French Ministry for Europe and Foreign Affairs. (2018). *Paris call for trust and security in cyberspace*. https://pariscall.international/en/call
Freyburg, T., & Garbe, L. (2018). Blocking the Bottleneck: Internet Shutdowns and Ownership at Election Times in Sub-Saharan Africa. *International Journal of Communication*, *12*, 3896–3916. http://www.ijoc.org/index.php/ijoc/article/download/8546/2464
Fruhlinger, J. (2017). Petya ransomware and NotPetya malware: What you need to know now. https://www.csoonline.com/article/3233210/ransomware/petya-ransomware-and-notpetya-malware-what-you-need-to-know-now.html
Fruhlinger, J. (2022). WannaCry explained: A perfect ransomware storm. https://www.csoonline.com/article/3227906/wannacry-explained-a-perfect-ransomware-storm.html
Fulghum, D. (2007). Why syria's air defenses failed to detect israelis. *Aviation Week & Space Technology*. https://cyber-peace.org/wp-content/uploads/2016/11/IMRA-Friday-October-5-2007-Why-Syrias-Air-Defenses-Failed-to-Detect-Israelis.pdf
Futter, A. (2020). What does cyber arms control look like ? Four principles for managing cyber risk GLOBAL SECURITY POLICY BRIEF. https://www.europeanleadershipnetwork.org/policy-brief/what-does-cyber-arms-control-look-like-four-principles-for-managing-cyber-risk/
Gais, A. (2019). The illusion of a "Great debate" about german security policy. A plea for more citizen participation. In G. Hellmann & D. Jacobi (Eds.), *The german white paper 2016 and the challenge of crafting security strategies* (pp. 77–87). Aspen Institute. https://www.fb03.uni-frankfurt.de/76345851/Band_Crafting_Security_Strategies_Aspen_englisch.pdf

Gatlan, S. (2020). *Microsoft: Emotet Took Down a Network by Overheating All Computers*. BleepingComputer. https://www.bleepingcomputer.com/news/security/microsoft-emotet-took-down-a-network-by-overheating-all-computers/

Gazette, S. (2016). Saudi embassy in Tehran attacked by protesters. *Saudi Gazette*. https://web.archive.org/web/20160206202830/http://saudigazette.com.sa/saudi-arabia/saudi-embassy-in-tehran-attacked-by-protesters/.

Geers, K. (2010). Cyber Weapons Convention. *Computer Law & Security Review*, *26*(5), 547–551. https://doi.org/10.1016/j.clsr.2010.07.005

Georgieva, I. (2020). The unexpected norm-setters: Intelligence agencies in cyberspace. *Contemporary Security Policy*, *41*(1), 33–54. https://doi.org/10.1080/13523260.2019.1677389

Gerlach, C. & Seitz, R. (2013). Economic impact of submarine cable disruptions, 96. https://www.apec.org/publications/2013/02/economic-impact-of-submarine-cable-disruptions

German Federal Parliament Defense Commitee. (2016). Wortprotokoll der 61. Sitzung des Verteidigungsausschußes des Deutschen Bundestages. www.bundestag.de/blob/417878/d8a5369a9df83e438814791a2881c5ef/protokoll-cyber-data.pdf

German Government. (2016). Weißbuch 2016 - zur sicherheitspolitik und zur zukunft der bundeswehr. https://www.bmvg.de/resource/blob/13708/015be272f8c0098f1537a491676bfc31/weissbuch2016-barrierefrei-data.pdf

German Government. (2021). On the Application of International Law in Cyberspace - Position Paper of the German Government, 17. https://www.auswaertiges-amt.de/blob/2446304/32e7b2498e10b74fb17204c54665bdf0/on-the-application-of-international-law-in-cyberspace-data.pdf

German Ministry for Interior Affairs. (2009). Nationale Strategie zum Schutz Kritischer Infrastrukturen (KRITIS-Strategie). https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/bevoelkerungsschutz/kritis.html

German Ministry of Defense. (2016). Abschlussbericht aufbaustab cyber- und informationsraum. http://docs.dpaq.de/11361-abschlussbericht_aufbaustab_cir.pdf

Giles, K., & Hartmann, K. (2019). "Silent Battle" Goes Loud: Entering a New Era of State-Avowed Cyber Conflict. *2019 11th International Conference on Cyber Conflict (CyCon)*, *900*, 1–13. https://doi.org/10.23919/CYCON.2019.8756713

Gillis, M. (2017). *Disarmament: A Basic Guide – Fourth Edition (2017) – UNODA*. https://www.un.org/disarmament/publications/basic-guide/disarmament-a-basic-guide-fourth-edition-2017/

Giraldo, J., Urbina, D., Cardenas, A. A., & Tippenhauer, N. O. (2019). Hide and Seek: An Architecture for Improving Attack-Visibility in Industrial Control Systems. In R. H. Deng, V. Gauthier-Umaña, M. Ochoa, & M. Yung (Eds.), *Applied Cryptography and Network Security* (pp. 175–195). Springer International Publishing. https://doi.org/10.1007/978-3-030-21568-2_9

Gisel, L., & Rodenhäuser, T. (2019, November 28). *Cyber operations and international humanitarian law: Five key points*. Humanitarian Law & Policy Blog. https://blogs.icrc.org/law-and-policy/2019/11/28/cyber-operations-ihl-five-key-points/

Gläser, J., & Laudel, G. (2010). *Experteninterviews und qualitative inhaltsanalyse als instrumente rekonstruierender untersuchungen* (4. Aufl). VS-Verl. https://link.springer.com/book/9783531172385

Global Commission on the Stability of Cyberspace. (2021). Statement on the interpretation of the norm on non-interference with the public core. https://hcss.nl/wp-content/uploads/2022/06/GCSC-statement-public-core_final.pdf.

Godehardt, N., & Voelsen, D. (2020). *NewIP – Grundstein für ein globales Internet nach chinesischen Vorstellungen?* Stiftung Wissenschaft und Politik (SWP). https://www.swp-berlin.org/publikation/newip-grundstein-fuer-ein-globales-internet-nach-chinesischen-vorstellungen

Goldblat, J. (2002). *Arms Control: The New Guide to Negotiations and Agreements*. Sage Publications. https://www.sipri.org/publications/2002/arms-control-new-guide-negotiations-and-agreements

Goldreich, O., Micali, S., & Wigderson, A. (1987). How to play ANY mental game. *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, 218–229. https://doi.org/10.1145/28395.28420

Government, U. (2016). National cyber security strategy 2016-2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/567242/national_cyber_security_strategy_2016.pdf

Granerud, A. O. (2010). *Identifying TLS abnormalities in Tor* [Master's thesis]. http://hdl.handle.net/11250/143950

Granick, J., & Fidler, M. (2014, January 15). *Changes to Export Control Arrangement Apply to Computer Exploits and More*. Changes to export control arrangement apply to computer exploits and more. https://www.justsecurity.org/5703/export-control-arrangement-apply-computer-exploits/

GReAT. (2017). WannaCry ransomware used in widespread attacks all over the world. *Securelist.Com*. https://securelist.com/wannacry-ransomware-used-in-widespread-attacks-all-over-the-world/78351/

Greenberg, A. (2018). Russian Hacker False Flags Work—Even After They're Exposed [magazine]. *Wired*. https://www.wired.com/story/russia-false-flag-hacks/

Greenberg, A. (2020). Russia's GRU Hackers Hit US Government and Energy Targets [magazine]. *Wired*. https://www.wired.com/story/russia-fancy-bear-us-hacking-campaign-government-energy/

Greenberg, A. (2023). Ukraine Suffered More Data-Wiping Malware in 2022 Than Anywhere, Ever [magazine]. *Wired*. https://www.wired.com/story/ukraine-russia-wiper-malware/

Grieco, J. M. (1988). Anarchy and the limits of cooperation: A realist critique of the newest liberal institutionalism. *International Organization*, *42*(3), 485–507. https://doi.org/10.1017/S0020818300027715

Grosswald, L. (2011). Cyberattack attribution matters under article 51 of the U.N. *Charter. Brooklyn Journal of International Law*, *36*(3). https://brooklynworks.brooklaw.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1124&context=bjil

Guardian, T. (2013). Obama tells intelligence chiefs to draw up cyber target list – full document text. *The Guardian*. https://www.theguardian.com/world/interactive/2013/jun/07/obama-cyber-directive-full-text

Guarnieri, C. (2015). Investigation report on the attacks on the IT of the parliamentary group DIE LINKE in the bundestag. https://cyber-peace.org/wp-content/uploads/2015/06/guarnieri-deutsch1.pdf

Guerrero-Saade, J. A., & Raiu, C. (2017). Walking in your enemy's shadow: When fourth-party collection becomes attribution hell. *Virus Bulletin Conference*. https://media.kasperskycontenthub.com/wp-content/uploads/sites/43/2018/03/07170728/Guerrero-Saade-Raiu-VB2017.pdf

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). XAI — Explainable artificial intelligence. *Science Robotics*, *4*(37), eaay7120. https://doi.org/10.1126/scirobotics.aay7120

Gupta, D., Mood, B., Feigenbaum, J., Butler, K., & Traynor, P. (2016). Using Intel Software Guard Extensions for Efficient Two-Party Secure Function Evaluation. In J. Clark, S. Meiklejohn, P. Y. Ryan, D. Wallach, M. Brenner, & K. Rohloff (Eds.), *Financial Cryptography and Data Security* (pp. 302–318). Springer. https://doi.org/10.1007/978-3-662-53357-4_20

Gupta, S. S. (2017). *Blockchain*. Indian Statistical Institute. https://www.isical.ac.in/~debrup/slides/Bitcoin.pdf

Gür, G., Bahtiyar, Ş., & Alagöz, F. (2015, January 1). Chapter 30 - Security analysis of computer networks: Key concepts and methodologies. In M. S. Obaidat, P. Nicopolitidis, & F. Zarai (Eds.), *Modeling and Simulation of Computer Networks and Systems* (pp. 861–898). Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-800887-4.00030-4

Hadaway, R., Hartling, E. R., Mehta, P., Hubbard, M., Evans, D., Berg, L., & Hinds, M. (2016, January 1). Submarine cable upgrades. In J. Chesnoy (Ed.), *Undersea Fiber Communication Systems (Second Edition)* (pp. 577–603). Academic Press. https://doi.org/10.1016/b978-0-12-804269-4.00016-7

Hansel, M., & Silomon, J. (2021). *Vertrauen stärken, neue Partner gewinnen – Deutschlands Beitrag für mehr Stabilität im Cyberraum - In Rüstungskontrolle für die nächste Bundesregierung. Ein Empfehlungsbericht* (IFSH Research Report (6).). https://ifsh.de/file/publication/Research_Report/006/Research_Report_006.pdf

Hansel, M., Mutschler, M., & Dickow, M. (2018). Taming cyber warfare: Lessons from preventive arms control. *Journal of Cyber Policy*, *3*(1), 44–60. https://doi.org/10.1080/23738871.2018.1462394

Hansel, M., & Nanni, S. (2018). Quantitative Rüstungsanalysen im Zeichen von Digitalisierung und Automatisierung. *Zeitschrift für Internationale Beziehungen*, *25*(1), 211–220. https://www.jstor.org/stable/26865059

Hao, Y., Jia, L., Wang, Y., & He, Z. (2021). Modelling cascading failures in networks with the harmonic closeness (C. Cherifi, Ed.). *PLOS ONE*, *16*(1), e0243801. https://doi.org/10.1371/journal.pone.0243801

Hao, Y., Wang, Y., Jia, L., & He, Z. (2020). Cascading failures in networks with the harmonic closeness under edge attack strategies. *Chaos, Solitons & Fractals*, *135*, 109772. https://doi.org/10.1016/j.chaos.2020.109772

Harknett, R. J., & Smeets, M. (2020). Cyber campaigns and strategic outcomes. *Journal of Strategic Studies*, *45*(4), 534–567. https://doi.org/10.1080/01402390.2020.1732354

Hatch, B. (2018). Defining a Class of Cyber Weapons as WMD: An Examination of the Merits. *Journal of Strategic Security*, *11*(1), 43–61. https://doi.org/10.5038/1944-0472.11.1.1657

Hau, B., Penrose, M., Hall, T., & Matias, B. (2016). M-Trends 2016. https://www2.fireeye.com/M-Trends-2016-EMEA-DE-LP.html

Hazay, C., & Venkitasubramaniam, M. (2017). Scalable Multi-party Private Set-Intersection. In S. Fehr (Ed.), *Public-Key Cryptography – PKC 2017* (pp. 175–203). Springer. https://doi.org/10.1007/978-3-662-54365-8_8

Healey, J. (2017). *Building a more defensible cyberspace*. Columbia School of International and Public Affairs. https://www.sipa.columbia.edu/global-research-impact/initiatives/cyber/nyctf/defensible-cyberspace

Healey, J. (2019). The implications of persistent (and permanent) engagement in cyberspace. *Journal of Cybersecurity*, *5*(1), 1–25. https://doi.org/10.1093/cybsec/tyz008

Heide, D. (2018, March 14). Will der Bund die Cybersicherheit erhöhen, muss er den Datenschutz opfern. http://www.handelsblatt.com/my/politik/deutschland/cybersecurity-will-der-bund-die-cybersicherheit-erhoehen-muss-er-den-datenschutz-opfern/21070060.html?ticket=ST-546642-G2ZE5mIzhcUXbO5Jbvle-ap2

Heller, M. (2018). CIA attributes NotPetya attacks to russian spy agency. http://searchsecurity.techtarget.com/news/450433303/CIA-attributes-NotPetya-attacks-to-Russian-spy-agency

Herpig, S. (2019). *Securing artificial intelligence*. Stiftung Neue Verantwortung. https://www.stiftung-nv.de/sites/default/files/securing_artificial_intelligence.pdf

Herpig, S. (2017). Cyber operations: Defending political IT-Infrastructures. A comparative problem analysis supported by the transatlantic cyber forum. *Stiftung Neue Verantwortung*. https://www.stiftung-nv.de/sites/default/files/tcf-defending_political_lt-infrastructures-problem_analysis.pdf

Herpig, S., Morgus, R., & Sheniak, A. (2020). Active cyber defense - A comparative study on US , Israeli and German approaches. *Konrad Adenauer Foundation*. https://www.kas.de/documents/263458/263507/Active+Cyber+Defense+-+A+comparative+study+on+US,+Israeli+and+German+approaches.pdf

Herpig, S., & Reinhold, T. (2018, October). Spotting the Bear: Credible Attribution and Russian Operations in Cyberspace. In N. Popescu & S. Secrieru (Eds.), *Hacks, leaks and disruptions: Russian cyber strategies* (pp. 33–42, Vol. 148). European Union Institute for Security Studies (EUISS). https://www.jstor.org/stable/resrep21140.7

Herr, T. (2014). PrEP: A Framework for Malware and Cyber Weapons. *Journal of Information Warfare*, *13*(1), 87–106. https://doi.org/10.2307/26487013

Hess, G. D. (2021). The Impact of a Regional Nuclear Conflict between India and Pakistan: Two Views. *Journal for Peace and Nuclear Disarmament*, *4*, 163–175. https://doi.org/10.1080/25751654.2021.1882772

Hinck, G. (2018). Wassenaar export controls on surveillance tools: New exemptions for vulnerability research. https://lawfareblog.com/wassenaar-export-controls-surveillance-tools-new-exemptions-vulnerability-research

Hollick, M., & Katzenbeisser, S. (2019). Resilient Critical Infrastructures. In C. Reuter (Ed.), *Information Technology for Peace and Security: IT Applications and Infrastructures in Conflicts, Crises, War, and Peace* (pp. 305–318). Springer Fachmedien. https://doi.org/10.1007/978-3-658-25652-4_14

Holtom, P., & Bromley, M. (2010). The international arms trade: Difficult to define, measure, and control. https://www.armscontrol.org/act/2010_07-08/holtom-bromley

Hoon, K., Yeo, K., Azam, S., Shunmugam, B., & Boer, F. (2018). Critical review of machine learning approaches to apply big data analytics in DDoS forensics. *2018 International Conference on Computer Communication and Informatics (ICCCI), 1–5. IEEE*, 1–5. https://doi.org/10.1109/ICCCI.2018.8441286

Hopkins, N. (2012). US and China engage in cyber war games [newspaper]. *The Guardian: Technology*. https://www.theguardian.com/technology/2012/apr/16/us-china-cyber-war-games

Hoppenstedt, M. (2022). Attack on Viasat: U.S. intelligence investigates cyberattack on satellite internet. *Spiegel Online*, *13*. https://www.spiegel.de/netzwelt/web/vias

at-nsa-untersucht-hacker-angriff-auf-satellitennetzwerk-a-caab89d0-7eac-444f-b488-b51369762749.

Huang, Y., Evans, D., & Katz, J. (2012). Private set intersection: Are garbled circuits better than custom protocols? *Symposium on Network and Distributed System Security (NDSS'12)*. https://www.cs.virginia.edu/~evans/pubs/ndss2012/psi.pdf

Hummelholm, A. (2019). Undersea optical cable network and cyber threats. *European Conference on Cyber Warfare and Security*. https://www.proquest.com/openview/7820d1f4cdd958cada85bdcd64ac6e68/1?cbl=396497&pq-origsite=gscholar&parentSessionId=klnuSIRfMW3O1JsodR72JCPWmy9lFu14nOLxVjjF5tg%3D

Hutchins, E., Cloppert, M., & Amin, R. (2011). Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. *6th International Conference on Information Warfare and Security, ICIW 2011*, 113–125. https://books.google.de/books?hl=de&lr=&id=oukNfumrXpcC&oi=fnd&pg=PA80&dq=Intelligence-driven+computer+network+defense+informed+by+analysis+of+adversary+campaigns+and+intrusion+kill+chains&ots=fcMO8vmT_h&sig=Q4fJJ_bHhQ3kCtsRZ5VpQsOMvdk&redir_esc=y#v=onepage&q=Intelligence-driven%20computer%20network%20defense%20informed%20by%20analysis%20of%20adversary%20campaigns%20and%20intrusion%20kill%20chains&f=false

Hyperledger. (2017). Hyperledger Composer Overview. https://www.hyperledger.org/wp-content/uploads/2017/05/Hyperledger-Composer-Overview.pdf

IAEA. (1961). The Agencys Safeguards. *International Atomic Energy Agency*. https://www.iaea.org/sites/default/files/publications/documents/infcircs/1961/infcirc26.pdf

IAEA. (2015a). Joint comprehensive plan of action. http://eeas.europa.eu/archives/docs/statements-eeas/docs/iran_agreement/iran_joint-comprehensive-plan-of-action_en.pdf

IAEA. (2015b). Verification and monitoring in the Islamic Republic of Iran in light of United Nations Security Council resolution 2231 (2015). https://www.iaea.org/sites/default/files/22/11/gov2022-62.pdf

IAEA. (2016). Iran and the IAEA: Verification and monitoring under the JCPOA. *International Atomic Energy Agency*. https://www.iaea.org/sites/default/files/publications/magazines/bulletin/bull57-2/5722627.pdf

Iansiti, M., & Lakhani, K. R. (2017). The truth about blockchain. *Harvard Busienss Review*. https://enterprisersproject.com/sites/default/files/the_truth_about_blockchain.pdf

IBM. (2021). Informix Servers: 12.10: Estimate the size and number of log files. https://www.ibm.com/docs/en/informix-servers/12.10?topic=files-estimate-size-number-log

ICRC. (1949). The geneva conventions of 12 august 1949. https://www.icrc.org/en/doc/assets/files/publications/icrc-002-0173.pdf

ICRC. (2022a). *ICRC proposes digital red cross/crescent emblem to signal protection in cyberspace*. ICRC proposes digital red cross/crescent emblem to signal protection in cyberspace. https://www.icrc.org/en/document/icrc-proposes-digital-red-crosscrescent-emblem-signal-protection-cyberspace

ICRC. (2022b). Russia-Ukraine international armed conflict: Immense damage to essential infrastructure will cause majorsuffering as winter looms. https://www.icrc

.org/en/document/russia-ukraine-international-armed-conflict-immense-damage-essential-infrastructure

Inbar, R., Omri, E., & Pinkas, B. (2018). Efficient Scalable Multiparty Private Set-Intersection via Garbled Bloom Filters. In D. Catalano & R. De Prisco (Eds.), *Security and Cryptography for Networks* (pp. 235–252). Springer International Publishing. https://doi.org/10.1007/978-3-319-98113-0_13

INF. (1988). Treaty between the united states of america and the union of soviet socialist republics on the elimination of their intermediate-range and shorter-range missiles (INF treaty. https://www.state.gov/t/avc/trty/102360.htm#text

Infrapedia. (2023). *Infrastucture map*. https://www.infrapedia.com/

iPRAW. (2019). *Focus on Human Control* (Vol. 5). https://nbn-resolving.org/urn:nbn:de:0168-ssoar-77410-6

Ipsos Public Affairs. (2018). Ipsos Poll Conducted for Reuters, Russia Poll 7.18.2018. https://www.ipsos.com/sites/default/files/ct/news/documents/2018-07/2018_reuters_tracking_-_russia_7_18_2018.pdf

ITU Publications. (2012). *Declarations and Reservations made at the end of the World Conference on International Telecommunications of the International Telecommunication Union (Dubai, 2012)*. Dubai. https://www.itu.int/dms_pub/itu-s/opb/conf/S-CONF-WCIT-2012-PDF-E.pdf

ITU Publications. (2018). Maximising availability of international connectivity in the Pacific. https://www.itu.int/en/ITU-D/Regulatory-Market/Documents/Infrastructure_portal/Maximising-availability-of-int-connectivity-in-the-pacific.pdf

ITU Publications. (2019a). *Small island developing states (SIDS) and ICTs*. https://doi.org/11.1002/pub/813cee7c-en

ITU Publications. (2019b). World Telecommunication/ICT indicators database. http://handle.itu.int/11.1002/pub/81377c7d-en

ITU Publications. (2020). *Internet Backbone fiber routes map*. https://bbmaps.itu.int/bbmaps/

Jervis, R. (1978). Cooperation Under the Security Dilemma. *World Politics*, *30*(2), 167–214. https://doi.org/10.2307/2009958

Ji-Young, K., Jong In, L., & Kyoung Gon, K. (2019). The All-Purpose Sword: North Korea's Cyber Operations and Strategies. *2019 11th International Conference on Cyber Conflict (CyCon)*, *900*, 1–20. https://doi.org/10.23919/CYCON.2019.8756954

Johnson, B., Caban, D., Krotofil, M., Scali, D., Brubaker, N., & Glyer, C. (2017). Attackers Deploy New ICS Attack Framework "TRITON" and Cause Operational Disruption to Critical Infrastructure. *Fireeye Threat Research*. https://www.fireeye.com/blog/threat-research/2017/12/attackers-deploy-new-ics-attack-framework-triton.html

Juniper. (2022). Junos ® OS interfaces user guide for security devices. https://www.juniper.net/documentation/us/en/software/junos/interfaces-security-devices/interfaces-security-devices.pdf

Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. *American Economic Review*, *93*(5), 1449–1475. https://doi.org/10.1257/000282803322655392

Kai, G. A., Tay, L., & Sharma, M. (2022, February 24). *GREAT-POWER OFFENSIVE CYBER CAMPAIGNS: Experiments in Strategy The International Institute for Strategic Studies GREAT-POWER OFFENSIVE CYBER CAMPAIGNS: Experiments in Strategy*. https://www.iiss.org/globalassets/media-library---co

ntent--migration/files/research-papers/2022/02/great-power-offensive-cyber-campaigns-experiments-in-strategy.pdf

Kales, D., Rechberger, C., Schneider, T., Senker, M., & Weinert, C. (2019). Mobile private contact discovery at scale. *USENIX Security'19*, 1447–1464. https://www.usenix.org/system/files/sec19-kales.pdf

Kamara, S., Mohassel, P., & Raykova, M. (2011). *Outsourcing Multi-Party Computation*. https://eprint.iacr.org/2011/272

Känzig, N., Meier, R., Gambazzi, L., Lenders, V., & Vanbever, L. (2019). Machine Learning-based Detection of C&C Channels with a Focus on the Locked Shields Cyber Defense Exercise. *2019 11th International Conference on Cyber Conflict (CyCon)*, *900*, 1–19. https://doi.org/10.23919/CYCON.2019.8756814

Karaganov, S., & Suslov, D. (2019). *The new understanding and ways to strengthen multilateral strategic stability* (Report.). Higher School of Economics, National Research University. Moscow. http://www.ewb.rs/upload/report_tass_eng.pdf

Kastelic, A., Biggio, G., Dalaqua, R. H., Kurosaki, M., Paoli, G. P., & Vignard, K. (2021). *International Cyber Operations: National Doctrines and Capabilities*. https://unidir.org/cyberdoctrines

Katina, P. F., Ariel Pinto, C., Bradley, J. M., & Hester, P. T. (2014). Interdependency-induced risk with applications to healthcare. *International Journal of Critical Infrastructure Protection*, *7*(1), 12–26. https://doi.org/10.1016/j.ijcip.2014.01.005

Katz, J., & Lindell, Y. (2014). *Introduction to modern cryptography*. Chapman and Hall/CRC. http://staff.ustc.edu.cn/~mfy/moderncrypto/reading%20materials/Introduction_to_Modern_Cryptography.pdf

Kaufhold, M.-A., Rupp, N., Reuter, C., & Habdank, M. (2020). Mitigating information overload in social media during conflicts and crises: Design and evaluation of a cross-platform alerting system. *Behaviour & Information Technology*, *39*(3), 319–342. https://doi.org/10.1080/0144929X.2019.1620334

Keller, M. (2020). MP-SPDZ: A versatile framework for multi-party computation. *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, 1575–1590. https://doi.org/10.1145/3372297.3417872

Keller, M., Orsini, E., & Scholl, P. (2016). MASCOT: Faster Malicious Arithmetic Secure Computation with Oblivious Transfer. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 830–842. https://doi.org/10.1145/2976749.2978357

Keohane, R. O. (1984). *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton University Press. https://www.jstor.org/stable/j.ctt7sq9s

Keohane, R. O., & Victor, D. G. (2011). The Regime Complex for Climate Change. *Perspectives on Politics*, *9*(1), 7–23. https://doi.org/10.1017/S1537592710004068

Kersting, K. (2018). Machine Learning and Artificial Intelligence: Two Fellow Travelers on the Quest for Intelligent Behavior in Machines. *Frontiers in Big Data*, *1*, 4. https://www.frontiersin.org/articles/10.3389/fdata.2018.00006

Kersting, K., & Meyer, U. (2018). From Big Data to Big Artificial Intelligence? *KI - Künstliche Intelligenz*, *32*(1), 3–8. https://doi.org/10.1007/s13218-017-0523-7

Khandelwal, S. (2018). Cryptocurrency mining malware infected over half-million PCs using NSA exploit. *The Hacker News*. http://thehackernews.com/2018/01/cryptocurrency-mining-malware.html

Kirat, D., & Vigna, G. (2015). MalGene: Automatic Extraction of Malware Analysis Evasion Signature. *Proceedings of the 22nd ACM SIGSAC Conference on*

Computer and Communications Security, 769–780. https://doi.org/10.1145/2810103.2813642

Kiss, Á., Liu, J., Schneider, T., Asokan, N., & Pinkas, B. (2017). Private Set Intersection for Unequal Set Sizes with Mobile Applications. *Proceedings on Privacy Enhancing Technologies*, 177–197. https://doi.org/10.1515/popets-2017-0044

Kitamura, Y., Lee, Y., Sakiyama, R., & Okamura, K. (2007). Experience with Restoration of Asia Pacific Network Failures from Taiwan Earthquake. *IEICE TRANSACTIONS on Communications*, *E90-B*(11), 3095–3103. https://search.ieice.org/bin/summary.php?id=e90-b_11_3095&category=B&year=2007&lang=E&abst=

Koch, R., & Golling, M. (2019). Silent Battles: Towards Unmasking Hidden Cyber Attack. *2019 11th International Conference on Cyber Conflict (CyCon)*, *900*, 1–20. https://doi.org/10.23919/CYCON.2019.8757146

Koeberl, P., Phegade, V., Rajan, A., Schneider, T., Schulz, S., & Zhdanova, M. (2015). Time to Rethink: Trust Brokerage Using Trusted Execution Environments. In M. Conti, M. Schunter, & I. Askoxylakis (Eds.), *Trust and Trustworthy Computing* (pp. 181–190). Springer International Publishing. https://doi.org/10.1007/978-3-319-22846-4_11

Kolesnikov, V., Matania, N., Pinkas, B., Rosulek, M., & Trieu, N. (2017). Practical Multi-party Private Set Intersection from Symmetric-Key Techniques. *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 1257–1272. https://doi.org/10.1145/3133956.3134065

Kolesnikov, V., Sadeghi, A.-R., & Schneider, T. (2009). Improved Garbled Circuit Building Blocks and Applications to Auctions and Computing Minima. In J. A. Garay, A. Miyaji, & A. Otsuka (Eds.), *Cryptology and Network Security* (pp. 1–20). Springer. https://doi.org/10.1007/978-3-642-10433-6_1

Kosseff, J. (2019). The Contours of 'Defend Forward' Under International Law. *2019 11th International Conference on Cyber Conflict (CyCon)*, *900*, 1–13. https://doi.org/10.23919/CYCON.2019.8757141

Kostyuk, N., & Zhukov, Y. M. (2019). Invisible Digital Front: Can Cyber Attacks Shape Battlefield Events? *Journal of Conflict Resolution*, *63*(2), 317–347. https://doi.org/10.1177/0022002717737138

Krasner, S. D. (1983). *International Regimes*. Cornell University Press. https://books.google.de/books?hl=de&lr=&id=WIYKBNM5zagC&oi=fnd&pg=PP8&dq=International+regimes+krasner&ots=pxApyNGYcf&sig=Qv9vMTOki9y9ggdYJYBHWdALBzA&redir_esc=y#v=onepage&q=International%20regimes%20krasner&f=false

Kraus, A. I., Frazer, O., Kirchhoff, L., Kyselova, T., Mason, S. J. A., & Federer, J. P. (2019). Dilemmas and Trade-Offs in Peacemaking: A Framework for Navigating Difficult Decisions. *Politics and Governance*, *7*(4), 331–342. https://doi.org/10.17645/pag.v7i4.2234

Kruse, J. (2015). *Qualitative Interviewforschung*. Juventa Verlag.

Kuckartz, U. (2018). Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung (5th). https://www.beltz.de/produkt_detailansicht/47304-qualitative-inhaltsanalyse-methoden-praxis-computerunterstuetzung.html

Küçük, K. A., Paverd, A., Martin, A., Asokan, N., Simpson, A., & Ankele, R. (2016). Exploring the use of Intel SGX for Secure Many-Party Applications. *Proceedings of the 1st Workshop on System Software for Trusted Execution*, 1–6. https://doi.org/10.1145/3007788.3007793

Kuehn, P., Bayer, M., Wendelborn, M., & Reuter, C. (2021). OVANA: An Approach to Analyze and Improve the Information Quality of Vulnerability Databases. *Proceedings of the 16th International Conference on Availability, Reliability and Security*, 1–11. https://doi.org/10.1145/3465481.3465744

Kuehn, P., Riebe, T., Apelt, L., Jansen, M., & Reuter, C. (2020). Sharing of cyber threat intelligence between states. *Sicherheit & Frieden*, *38*(1), 22–28. https://doi.org/10.5771/0175-274X-2020-1-22

Kühn, U. (2014). *Applying Insights Gained from Traditional TCBMs to Cyberspace*. S. Rajaratnam School of International Studies. http://www.jstor.org/stable/resrep05892.11

Kumar, R. S. S., Wicker, A., & Swann, M. (2017, September 20). *Practical Machine Learning for Cloud Intrusion Detection: Challenges and the Way Forward*. https://doi.org/10.48550/arXiv.1709.07095

Kumar, S., & Carley, K. M. (2016). Understanding DDoS cyber-attacks using social media analytics. *2016 IEEE Conference on Intelligence and Security Informatics (ISI)*, 231–236. https://doi.org/10.1109/ISI.2016.7745480

Kuntke, F., Linsner, S., Steinbrink, E., Franken, J., & Reuter, C. (2022). Resilience in Agriculture: Communication and Energy Infrastructure Dependencies of German Farmers. *International Journal of Disaster Risk Science*, *13*(2), 214–229. https://doi.org/10.1007/s13753-022-00404-7

Kütt, M., Göttsche, M., & Glaser, A. (2018). Information barrier experimental: Toward a trusted and open-source computing platform for nuclear warhead verification. *Measurement*, *114*, 185–190. https://doi.org/10.1016/j.measurement.2017.09.014

Langner, R. (2013). *To Kill a Centrifuge | Detailed Stuxnet Analysis | Langner*. OTbase by Langner. https://www.langner.com/to-kill-a-centrifuge/

Latora, V., & Marchiori, M. (2003). Economic small-world behavior in weighted networks. *The European Physical Journal B - Condensed Matter and Complex Systems*, *32*(2), 249–263. https://doi.org/10.1140/epjb/e2003-00095-5

Latora, V., & Marchiori, M. (2001). Efficient Behavior of Small-World Networks. *Physical Review Letters*, *87*(19), 198701. https://doi.org/10.1103/PhysRevLett.87.198701

Lavrenovs, A. (2019). Towards Measuring Global DDoS Attack Capacity. *2019 11th International Conference on Cyber Conflict (CyCon)*, *900*, 1–15. https://doi.org/10.23919/CYCON.2019.8756851

Lee, M.-J. (2016, January 19). *Igraphinshiny: Use 'shiny' to Demo 'igraph'* (Version 0.1). https://CRAN.R-project.org/package=igraphinshiny

Lemay, A., Calvet, J., Menet, F., & Fernandez, J. M. (2018). Survey of publicly available reports on advanced persistent threat actors. *Computers & Security*, *72*, 26–59. https://doi.org/10.1016/j.cose.2017.08.005

Lewis, J. (2010a). Cyberwarfare and its impact on international security. In *UNODA occasional paper 19*. https://www.armscontrol.org/act/2010-06/multilateral-agreements-constrain-cyberconflict

Lewis, J. (2010b). *Multilateral agreements to constrain cyberconflict*. Arms Control Today. https://www.armscontrol.org/act/2010-06/multilateral-agreements-constrain-cyberconflict

Lewis, J. (2013). Conflict and negotiation in cyberspace. https://www.csis.org/analysis/conflict-and-negotiation-cyberspace

Libicki, M. C. (2009, September 10). *Cyberdeterrence and Cyberwar*. RAND Corporation. https://www.rand.org/pubs/monographs/MG877.html

Lilly, B., Ablon, L., Hodgson, Q. E., & Moore, A. S. (2019). Applying Indications and Warning Frameworks to Cyber Incidents. *2019 11th International Conference on Cyber Conflict (CyCon)*, *900*, 1–21. https://doi.org/10.23919/CYCON.2019.8756949

Lin, H. (2011). On Attribution and Defense. *International Conference on Challenges in Cybersecurity – Risks, Strategies, and Confidence-building*.

Lindell, Y., Smart, N. P., & Soria-Vazquez, E. (2016). More Efficient Constant-Round Multi-party Computation from BMR and SHE. In M. Hirt & A. Smith (Eds.), *Theory of Cryptography* (pp. 554–581). Springer. https://doi.org/10.1007/978-3-662-53641-4_21

Lindsay, J. R. (2015). Tipping the scales: The attribution problem and the feasibility of deterrence against cyberattack. *Journal of Cybersecurity*, *1*(1), 53–67. https://doi.org/10.1093/cybsec/tyv003

Litwak, R., & King, M. (2015). Arms control in cyberspace? In *Wilson briefs*. https://www.wilsoncenter.org/sites/default/files/media/documents/publication/arms_control_in_cyberspace.pdf

Liu, S., Bischof, Z. S., Madan, I., Chan, P. K., & Bustamante, F. E. (2020). Out of sight, not out of mind: A user-view on the criticality of the submarine cable network. *Proceedings of the ACM Internet Measurement Conference*, 194–200. https://doi.org/10.1145/3419394.3423633

Lück, N. (2019). *Lernende Künstliche Intelligenz in der Rüstungskontrolle* (Vol. 4). Hessische Stiftung Friedens- und Konfliktforschung. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-64358-3

Luhmann, N. (1984). *Soziale Systeme: Grundriss einer allgemeinen Theorie*. Suhrkamp. https://www.suhrkamp.de/buch/niklas-luhmann-soziale-systeme-t-9783518282663

Lunden, K., Kapellmann Zafra, D., & Brubaker, N. (2021). Crimes of opportunity: Increasing frequency of low sophistication operational technology compromises. *Mandiant*. https://www.mandiant.com/resources/increasing-low-sophistication-operational-technology-compromises

Lyu, J. (2019). What are china's cyber capabilities and intentions? https://carnegieendowment.org/2019/04/01/what-are-china-s-cyber-capabilities-and-intentions-pub-78734.

Maathuis, C., Pieters, W., & Den Berg, J. V. (2016). Cyber weapons: A profiling framework. *2016 International Conference on Cyber Conflict (CyCon U.S.)*, 1–8. https://doi.org/10.1109/CYCONUS.2016.7836621

Maersk. (2017). Maersk improves underlying profit and grows revenue in first half of the year. https://cyber-peace.org/wp-content/uploads/2017/07/A.P.pdf

Mahlknecht, G. (2016). *Greg's Cable Map*. https://cablemap.info/_default.aspx

Malwarebytes Blog. (2021). *Trojan.Emotet*. Malwarebytes Labs. https://blog.malwarebytes.com/detections/trojan-emotet/

Mandiant Corporation. (2013). APT1 - Exposing one of China's cyber espionage units. https://www.fireeye.com/content/dam/fireeye-www/services/pdfs/mandiant-apt1-report.pdf

Mandiant Corporation. (2017). M-trends 2017: A view from the front lines. https://www2.fireeye.com/rs/848-DID-242/images/RPT-M-Trends-2017.pdf

Mandiant Corporation. (2020). *M-Trends 2020*. Mandiant. https://mandiant.widen.net/s/5pwlhgqt5t/m-trends-2020-report

Manthey, F., & Fleischer, J. (2023, March 31). *Bundeskabinett beschließt Cyberagentur*. https://www.bmvg.de/de/aktuelles/bundeskabinett-beschliesst-cyberagentur-27392

Marcus, J. (2019). Why Saudi Arabia and Iran are bitter rivals [newspaper]. *BBC News: Middle East*. https://www.bbc.com/news/world-middle-east-42008809

Martin, B., Brown, M., Paller, A., Kirby, D., & Christey, S. (2011). *CWE/SANS Top 25 Most Dangerous Software Errors*. MITRE Cooperation. https://cwe.mitre.org/top25/archive/2011/2011_cwe_sans_top25.pdf

Marzo, J. L., Calle, E., Cosgaya, S. G., Rueda, D., & Manosa, A. (2018). On selecting the relevant metrics of network robustness. *2018 10th International Workshop on Resilient Networks Design and Modeling (RNDM)*. https://doi.org/10.1109/rndm.2018.8489809

Maslov, G., & Ustinov, R. (2020). *IAEA Verification Activities: A Tool of Building Trust or Building Up Pressure?* Russia in Global Affairs. https://eng.globalaffairs.ru/articles/iaea-building-trust/

Mathews, L. (2017). NotPetya ransomware attack cost shipping giant maersk over 200 million [newspaper]. *Forbes*. https://www.forbes.com/sites/leemathews/2017/08/16/notpetya-ransomware-attack-cost-shipping-giant-maersk-over-200-million/#a6fe20f4f9ae.

Mauldin, A. (2017, May 3). *Cable Breakage: When and How Cables Go Down*. https://blog.telegeography.com/what-happens-when-submarine-cables-break

Maurer, T. (2015). Cyber proxies and the crisis in ukraine. In K. Geers (Ed.), *Cyber war in perspective: Russian aggression against ukraine* (pp. 79–86). NATO CCD COE Publications. https://ccdcoe.org/uploads/2018/10/Ch09_CyberWarinPerspective_Maurer.pdf

Maxey, L. (2018). False flags in cyberspace: Targeting public opinion and political will. *The Cipher Brief*. https://www.thecipherbrief.com/false-flags-cyberspace-targeting-public-opinion-political-will

Maybaum, M., & Tölle, J. (2016). Arms control in cyberspace - architecture for a trust-based implementation framework based on conventional arms control methods. *2016 8th International Conference on Cyber Conflict (CyCon)*, 159–173. https://doi.org/10.1109/CYCON.2016.7529433

Mayer, H. O. (2012, December 4). Interview und schriftliche Befragung: Grundlagen und Methoden empirischer Sozialforschung. In *Interview und schriftliche Befragung* (6th ed.). Oldenbourg Wissenschaftsverlag. https://doi.org/10.1524/9783486717624

Mayring, P. (2015). Qualitative Inhaltsanalyse. Grundlagen und Techniken (12.). https://www.beltz.de/fachmedien/paedagogik/produkte/details/48632-qualitative-inhaltsanalyse.html

Mayring, P., & Fenzl, T. (2014). Qualitative Inhaltsanalyse. In N. Baur & J. Blasius (Eds.), *Handbuch Methoden der empirischen Sozialforschung* (pp. 543–556). Springer Fachmedien. https://doi.org/10.1007/978-3-531-18939-0_38

McCulloh, I., & Perrone, A. (2017). R Packages for Social Network Analysis. In R. Alhajj & J. Rokne (Eds.), *Encyclopedia of Social Network Analysis and Mining* (pp. 1–18). Springer. https://doi.org/10.1007/978-1-4614-7163-9_110209-1

McDonald, G., O Murchu, L., Doherty, S., & Chien, E. (2013). Stuxnet 0.5: The Missing Link. *Symantec*, 18. https://docs.broadcom.com/doc/stuxnet-missing-link-13-en

Meier, O. (2020). Yes, we can? Europäische Antworten auf die Krise der Rüstungskontrolle. *Policy Brief*, *7*, 1–8. https://ifsh.de/file/publication/Policy_Brief/20_07_Policy_Brief.pdf

Mele, S. (2013, June 7). *Cyber-Weapons: Legal and Strategic Aspects (Version 2.0)*. https://doi.org/10.2139/ssrn.2518212

Meuser, M., & Nagel, U. (2009). Das Experteninterview — konzeptionelle Grundlagen und methodische Anlage. In S. Pickel, G. Pickel, H.-J. Lauth, & D. Jahn (Eds.), *Methoden der vergleichenden Politik- und Sozialwissenschaft: Neue Entwicklungen und Anwendungen* (pp. 465–479). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91826-6_23

Meyer, P. (2011). Cyber-Security Through Arms Control. *The RUSI Journal*, *156*(2), 22–27. https://doi.org/10.1080/03071847.2011.576471

Microsoft. (2022). Special report: Ukraine - an overview of russia's cyberattack activity in ukraine. *Microsoft April*, *27*. https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE4Vwwd

Milch, R. S., Pernice, I., Romanosky, S., von Lewinski, K., Shackelford, S. J., Rosenzweig, P., Christakis, T., Swire, P., Healey, J., Herpig, S., Pohle, J., Zatko, S., & Wenger, E. (2020). Building Common Approaches for Cybersecurity and Privacy in a Globalized World. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3508933

Miller, C. (2017, July 6). Apple's bug bounty program faltering due to low payouts to researchers, new report claims. https://9to5mac.com/2017/07/06/apple-bug-bounty-program-payouts/.

Miller, S., Brubaker, N., Zafra, D. K., & Caban, D. (2019). TRITON actor TTP profile, custom attack tools, detections, and ATT&CK mapping. https://www.fireeye.com/blog/threat-research/2019/04/triton-actor-ttp-profile-custom-attack-tools-detections.html

Mimoso, M. (2017). New petya distribution vectors bubbling to surface. *Threatpost.com*. https://threatpost.com/new-petya-distribution-vectors-bubbling-to-surface/126577/

Minárik, T. & Stinissen, Jan. (2014). From active cyber defence to responsive cyber defence: A way for states to defend themselves – legal implications [magazine]. *NATO Legal Gazette*, (35). https://www.act.nato.int/images/stories/media/doclibrary/legal_gazette_35.pdf

Mirhoseini, A., Sadeghi, A.-R., & Koushanfar, F. (2016). CryptoML: Secure outsourcing of big data machine learning applications. *2016 IEEE International Symposium on Hardware Oriented Security and Trust (HOST)*, 149–154. https://doi.org/10.1109/HST.2016.7495574

MITRE Cooperation. (2019). CWE - Research Concept. https://cwe.mitre.org/data/definitions/1000.html

MITRE Cooperation. (2023). *Common vulnerabilities and exposures*. https://cve.mitre.org/

Modderkolk, H. (2018, January 25). Dutch agencies provide crucial intel about russia's interference in US elections. In *De volksrant*. https://www.volkskrant.nl/media/dutch-agencies-provide-crucial-intel-about-russia-s-interference-in-us-elections~a4561913/

Mohassel, P., & Rindal, P. (2018). ABY ^{3}: A Mixed Protocol Framework for Machine Learning. *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, 35–52. https://doi.org/10.1145/3243734.3243760

Mohurle, S., & Patil, M. (2017). A brief study of wannacry threat: Ransomware attack 2017. *International Journal of Advanced Research in Computer Science (IJARCS*, *8*(5), 1938–1940. https://doi.org/10.26483/ijarcs.v8i5.4021

Mozilla. (2017). Security Bug Bounty Program. https://www.mozilla.org/en-US/security/bug-bounty/

Mulazzani, F., & Sarcia, S. A. (2011). Cyber security on military deployed networks - A Case Study on RealInformation Leakage. In E. C. Czosseck, E. Tyugu, & T. Wingfield (Eds.), *2011 3rd international conference on cyber conflict* (pp. 13–27). NATO. https://ccdcoe.org/uploads/2018/10/CyberSecurityOnMilitaryDeployedNetworks-Mulazzani-Sarcia.pdf

Müller, H. (2005). Multilaterale Abrüstung in der Krise. *Vereinte Nationen*, *2*, 41–45. https://doi.org/10.5771/0042-384X-2005-2

Müller, H., & Schörnig, N. (2006). *Rüstungsdynamik und Rüstungskontrolle: Eine exemplarische Einführung in die internationalen Beziehungen (Aussenpolitik und internationale Ordnung*. Nomos. https://www.nomos-shop.de/nomos/titel/ruestungsdynamik-und-ruestungskontrolle-id-77492/

Müller, H. (1982). Compliance politics: A critical analysis of multilateral arms control treaty enforcement. *The Nonproliferation Review*. https://www.nonproliferation.org/wp-content/uploads/npr/72muell.pdf

Muneez, M., Vinesh, T., & Nai-Shyan, L. (2017). Protection of submarine optical fibre cables on the coral reefs of the Maldives. *Underwater Technology*, *34*(4), 149–156. https://doi.org/10.3723/ut.34.149

Munich Security Conference. (2019). Münchner Sicherheitsreport 2019. The great puzzle: Who will pick up the pieces? MSC. https://doi.org/10.47342/RYTY8045

Murphy, M. (2019). *The cybersecurity protection of peacetime organizations: Comprehensive test ban treaty organization* [M.S.]. https://www.proquest.com/docview/2226137950/abstract/AC60DC5BF2CE4F6DPQ/1

Nadeem, S. (2019). *If we lived in a Bitcoin future, how big would the blockchain have to be?* Hackernoon. https://hackernoon.com/if-we-lived-in-a-bitcoin-future-how-big-would-the-blockchain-have-to-be-bd07b282416f

Nagurney, A., & Qiang, Q. (2007). A network efficiency measure with application to critical infrastructure networks. *Journal of Global Optimization*, *40*(1-3), 261–275. https://doi.org/10.1007/s10898-007-9198-1

Nagy, L. (2019). *VB2019 paper : Exploring Emotet , an elaborate everyday enigma - A brief history of Emotet*. Virus Bulletin. https://www.virusbulletin.com/virusbulletin/2019/10/vb2019-paper-exploring-emotet-elaborate-everyday-enigma/

Nair, S. (2022, March 1). *Belarus hackers attack train systems to disrupt Russian troops*. Railway Technology. https://www.railway-technology.com/news/belarus-hackers-attack-train-systems/

Nakashima, E., & Mufson, S. (2015). The U.S. and China agree not to conduct economic espionage in cyberspace [newspaper]. *Washington Post: National Security*. https://www.washingtonpost.com/world/national-security/the-us-and-china-agree-not-to-conduct-economic-espionage-in-cyberspace/2015/09/25/1c03f4b8-63a2-11e5-8e9e-dce8a2a2a679_story.html

Nakashima, E., & Timberg, C. (2017). NSA officials worried about the day its potent hacking tool would get loose. Then it did. [newspaper]. *Washington Post: Technology*. https://www.washingtonpost.com/business/technology/nsa-officials-worried-about-the-day-its-potent-hacking-tool-would-get-loose-then-it-did/2017/05/16/50670b16-3978-11e7-a058-ddbb23c75d82_story.html

Nakashima, E., & Warrick, J. (2012). Stuxnet was work of U.S. and Israeli experts, officials say [newspaper]. *Washington Post: National Security*, 1–4. https://www.washingtonpost.com/world/national-security/stuxnet-was-work-of-us-and-israeli-experts-officials-say/2012/06/01/gJQAlnEy6U_story.html

Nassauer, O. (2008). *20 Jahre nach dem INF-Vertrag Rüstungskontrolle ohne Zukunft?* (BITS Research Report No. 1)). http://www.bits.de/public/researchreport/rr08-1-1.htm

NATO. (2016). Warsaw summit communiqué issued by the heads of state and government participating in the meeting of the north atlantic council in warsaw 8-9. http://www.nato.int/cps/en/natohq/official_texts_133169.htm

NATO. (2023, February 27). *Arms control, disarmament and non-proliferation in NATO*. nato.int. https://www.nato.int/cps/en/natohq/topics_48895.htm

NCCIC. (2018). Malware Analysis MAR-17-352-01 HatMan — Safety System Targeted Malware (Update A), 1–24. https://www.cisa.gov/sites/default/files/documents/MAR-17-352-01%20HatMan%20-%20Safety%20System%20Targeted%20Malware%20(Update%20A)_S508C.PDF

NCCIC. (2019). Malware Analysis MAR-17-352-01 HatMan — Safety System Targeted Malware (Update B). *Cybersecurity and Infrastructure Security Agency*, 1–23. https://www.cisa.gov/sites/default/files/documents/MAR-17-352-01%20HatMan%20-%20Safety%20System%20Targeted%20Malware%20%28Update%20B%29.pdf

Neuneck, G., & Mölling, C. (2001). Präventive Rüstungskontrolle und Information Warfare. In *Rüstungskontrolle im Cyberspace. Perspektiven der Friedenspolitik im Zeitalter von Computerattacken* (pp. 47–53). Dokumentation einer Internationalen Konferenz der Heinrich-Böll-Stiftung am.

Neuneck, G. (2017). 60 Jahre nuklearer Prometheus oder Sisyphos? *Vereinte Nationen: German Review on the United Nations*, *65*(4), 170–176. https://www.jstor.org/stable/48551110

Neuneck, G. (2018). Nukleare Rüstungskontrolle: Stand und zentrale Herausforderungen. *Zeitschrift für Außen- und Sicherheitspolitik*, *11*(4), 581–601. https://doi.org/10.1007/s12399-018-0742-5

NEW Start. (2010). Treaty between the united states of america and the russian federation on measures for the further reduction and limitation of strategic offensive arms. https://www.nti.org/wp-content/uploads/2021/09/aptnewstart.pdf

Newbill, C. (2019). Defining Critical Infrastructure for a Global Application. *26 Indiana J. Global Legal Studies 761 (2019)*, *26*(2). https://www.repository.law.indiana.edu/ijgls/vol26/iss2/11

Nguyen, Q., Pham, H. D., Cassi, D., & Bellingeri, M. (2019). Conditional attack strategy for real-world complex networks. *Physica A: Statistical Mechanics and its Applications*, *530*, 121561. https://doi.org/10.1016/j.physa.2019.121561

Ni, X., He, D., Chan, S., & Ahmad, F. (2016). Network Anomaly Detection Using Unsupervised Feature Selection and Density Peak Clustering. In M. Manulis, A.-R. Sadeghi, & S. Schneider (Eds.), *Applied Cryptography and Network Security* (pp. 212–227). Springer International Publishing. https://doi.org/10.1007/978-3-319-39555-5_12

Nielsen, J. B., Nordholt, P. S., Orlandi, C., & Burra, S. S. (2012). A New Approach to Practical Active-Secure Two-Party Computation. In R. Safavi-Naini & R. Canetti (Eds.), *Advances in Cryptology – CRYPTO 2012* (pp. 681–700). Springer. https://doi.org/10.1007/978-3-642-32009-5_40

NIST. (2012). Nist special publication 800-30 revision 1 - guide for conducting risk assessments. *NIST Guide for Conducting Risk Assessments*, (September). https://csrc.nist.gov/publications/detail/sp/800-30/rev-1/final

NIST. (2015). NIST policy on hash functions. https://csrc.nist.gov/projects/hash-functions/nist-policy-on-hash-functions

Noguchi, M., & Ueda, H. (2017). An analysis of the actual status of recent cyberattacks on critical infrastructures. *NEC Technical Journal*, *12*(2), 19–24. https://www.nec.com/en/global/techrep/journal/g17/n02/170204.html

NPT. (1970). Treaty on the non-proliferation of nuclear weapons. https://www.iaea.org/sites/default/files/publications/documents/infcircs/1970/infcirc140.pdf

Oberhaus, D. (2018, January 30). *Cryptocurrency Mining Malware That Uses an NSA Exploit Is On the Rise*. Vice. https://www.vice.com/en/article/yw5yp7/monero-mining-wannamine-wannacry-nsa

O'Connor, F. (2017). NotPetya's fiscal impact revised: 892.5 million and growing. https://www.cybereason.com/blog/blog-notpetyas-fiscal-impact-revised-892-5-million-and-growing

Oehlers, M., & Fabian, B. (2021). Graph Metrics for Network Robustness—A Survey. *Mathematics*, *9*(8), 895. https://doi.org/10.3390/math9080895

Oldham, S., Fulcher, B., Parkes, L., Arnatkevičiūtė, A., Suo, C., & Fornito, A. (2019). Consistency and differences between centrality measures across distinct classes of networks (S. Hayasaka, Ed.). *PLOS ONE*, *14*(7), e0220061. https://doi.org/10.1371/journal.pone.0220061

Olszewski, B. (2018). Advanced persistent threats as a manifestation of states' military activity in cyber space. *Scientific Journal of the Military University of Land Forces*, *189*(3), 57–71. https://doi.org/10.5604/01.3001.0012.6227

O'Malley, S. (2019). Assessing threats to south korea's undersea communications cable infrastructure. *The Korean Journal of International Studies*, *17*(3), 385–414. https://doi.org/10.14731/kjis.2019.12.17.3.385

Omer, M., Nilchiani, R., & Mostashari, A. (2009). Measuring the Resilience of the Trans-Oceanic Telecommunication Cable System. *IEEE Systems Journal*, *3*(3), 295–303. https://doi.org/10.1109/JSYST.2009.2022570

One, A. (1996). Smashing the stack for fun and profit. *Phrack magazine*, *7*(49), 14–16. https://inst.eecs.berkeley.edu/~cs161/fa08/papers/stack_smashing.pdf

OpenSSL. (2020). Openssl-dgst, dgst - perform digest operations. https://www.openssl.org/docs/man1.1.1/man1/openssl-dgst.html

Orye, E., & Maennel, O. M. (2019). Recommendations for Enhancing the Results of Cyber Effects. *2019 11th International Conference on Cyber Conflict (CyCon)*, *900*, 1–19. https://doi.org/10.23919/CYCON.2019.8756649

Osanaiye, O., Choo, K.-K. R., & Dlodlo, M. (2016). Distributed denial of service (DDoS) resilience in cloud: Review and conceptual cloud DDoS mitigation framework. *Journal of Network and Computer Applications*, *67*, 147–165. https://doi.org/10.1016/j.jnca.2016.01.001

OSCE. (1986). Document of the Stockholm Conference, 1986. *Document of the Stockholm Conference on Confidence- and Security-Building Measures and Disarmament in Europe Convened in Accordance with the Relevant Provisions of the Concluding Document of the Madrid Meeting of the Conference on Security and Co-Operation*, 22. https://www.osce.org/fsc/41238?download=true

OSCE. (1992). Treaty on Open Skies. https://www.osce.org/files/f/documents/1/5/14127.pdf

OSCE. (2011). Vienna Documentt 2011 on confidence- and security-building measures. https://www.osce.org/fsc/86597?download=true#page=1&zoom=auto,-276,842

OSCE. (2013). Initial set of OSCE confidence-building measures to reduce the risks of conflict stemming from the use of information and communication technologies. http://www.osce.org/pc/109168?download=true

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press. https://doi.org/10.1017/CBO9780511807763

Paganini, P. (2022a). Non state actors in cyberspace: An'Attempt to a taxonomic classification, role, impactand relations with a state's socio-economic structure. In *Commentary, center for cyber security and international relations studies, università degli studi di firenze*. https://www.cssii.unifi.it/upload/sub/Pubblicazioni/2022_Paganini_Pierluigi.pdf

Paganini, P. (2022b). Ukraine: Volunteer IT Army is going to hit tens of Russian targets from this list. *Security Affairs*. https://securityaffairs.co/wordpress/128464/cyber-warfare-2/ukraine-volunteer-it-army-targets-russia.html.

Palmer, D. (2017). *Petya ransomware attack: How many victims are there really?* NDNet. http://www.zdnet.com/article/petya-ransomware-attack-how-many-victims-are-there-really/.

Palmer-Felgate, A., & Booi, P. (2016). How resilient is the global submarine cable network. *How Resilient Is the Global Submarine Cable Network*, 1–7.

Palmer-Felgate, A., Irvine, N., Ratcliffe, S., & Bah, S. S. (2013). Marine maintenance in the zones: A global comparison of repair commencement times. *Suboptic Conference from Ocean to Cloud*, 22–25. https://minz.org.nz/i/2018-challenges/Marine-maintenance-in-the-zones.pdf

Parliament, E. (2018). Adopted text: Control of exports, transfer, brokering, technical assistance and transit of dual-use item. https://doi.org/10.1080/00344897208656356

Patra, A., & Suresh, A. (2020). BLAZE: Blazing Fast Privacy-Preserving Machine Learning. *Proceedings 2020 Network and Distributed System Security Symposium*. https://doi.org/10.14722/ndss.2020.24202

Pawlak, P. (2016). Confidence-building measures in cyberspace: Current debates and trends. In A.-M. Osula & H. Rögias (Eds.), *International cyber norms: Legal, policy & industry perspectives*. NATO CCD COE Publications. https://ccdcoe.org/uploads/2018/10/InternationalCyberNorms_Ch7.pdf

PCriskcom. (2017). Another banking trojan leverages EternalBlue. https://www.pcrisk.com/internet-threat-news/11729-another-banking-trojan-leverages-eternalblue.

Perkovich, G., & Hoffman, W. (2019). From Cyber Swords to Plowshares, 1–9. https://carnegieendowment.org/2019/10/14/from-cyber-swords-to-plowshares-pub-80035

Perkovich, G., & Levite, A. E. (2017). From Understanding Cyber Conflict: Fourteen Analogies. *Georgetown University Press*. https://carnegieendowment.org/2017/10/16/understanding-cyber-conflict-14-analogies-pub-72689

Perlroth, N. (2015). HackerOne Connects Hackers With Companies, and Hopes for a Win-Win [newspaper]. *The New York Times: Technology*. https://www.nytimes.com/2015/06/08/technology/hackerone-connects-hackers-with-companies-and-hopes-for-a-win-win.html

Perlroth, N., & Krauss, C. (2018). A Cyberattack in Saudi Arabia Had a Deadly Goal. Experts Fear Another Try. [newspaper]. *The New York Times: Technology*. https://www.nytimes.com/2018/03/15/technology/saudi-arabia-hacks-cyberattacks.html

Perry, L., Shapira, B., & Puzis, R. (2019). NO-DOUBT: Attack Attribution Based On Threat Intelligence Reports. *2019 IEEE International Conference on Intelligence and Security Informatics (ISI)*, 80–85. https://doi.org/10.1109/ISI.2019.8823152

Peterson, A. (2021). The Sony Pictures hack, explained [newspaper]. *Washington Post*. https://www.washingtonpost.com/news/the-switch/wp/2014/12/18/the-sony-pictures-hack-explained/

Petrik, R., Arik, B., & Smith, J. M. (2018). Towards Architecture and OS-Independent Malware Detection via Memory Forensics. *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, 2267–2269. https://doi.org/10.1145/3243734.3278527

Pimenta Rodrigues, G. A., De Oliveira Albuquerque, R., Gomes de Deus, F. E., De Sousa Jr., R. T., De Oliveira Júnior, G. A., García Villalba, L. J., & Kim, T.-H. (2017). Cybersecurity and Network Forensics: Analysis of Malicious Traffic towards a Honeynet with Deep Packet Inspection. *Applied Sciences*, *7*(10), 1082. https://doi.org/10.3390/app7101082

Pinkas, B., Rosulek, M., Trieu, N., & Yanai, A. (2020). PSI from PaXoS: Fast, Malicious Private Set Intersection. In A. Canteaut & Y. Ishai (Eds.), *Advances in Cryptology – EUROCRYPT 2020* (pp. 739–767). Springer International Publishing. https://doi.org/10.1007/978-3-030-45724-2_25

Pinkas, B., Schneider, T., & Zohner, M. (2018). Scalable Private Set Intersection Based on OT Extension. *ACM Transactions on Privacy and Security*, *21*(2), 7:1–7:35. https://doi.org/10.1145/3154794

Pitrelli, M. (2022). Hacktivist group anonymous is using six top techniques to 'embarrass' russia [newspaper]. *CNBC: Tech*. https://www.cnbc.com/2022/07/28/how-is-anonymous-attacking-russia-the-top-six-ways-ranked-.html

Purdon, I., & Erturk, E. (2017). Perspectives of Blockchain Technology, its Relation to the Cloud and its Potential Role in Computer Science Education. *Engineering, Technology & Applied Science Research*, *7*(6), 2340–2344. https://doi.org/10.48084/etasr.1629

Putz, B., Menges, F., & Pernul, G. (2019). A secure and auditable logging infrastructure based on a permissioned blockchain. *Computers & Security*, *87*, 101602. https://doi.org/10.1016/j.cose.2019.101602

Qiu, W. (2020). Southern cross NEXT cable system overview. https://www.submarinenetworks.com/en/systems/trans-pacific/southern-cross-next/southern-cross-next-cable-system-overview

Reinhold, T. (2014a, April 22). *Die neuen digitalen Waffenhändler?* cyber-peace.org. https://cyber-peace.org/2014/04/22/die-neuen-digitalen-waffenhaendler/

Reinhold, T. (2018a). *Maßnahmen für den Cyberpeace*. cyber-peace.org. https://cyber-peace.org/cyberpeace-cyberwar/masnahmen-fur-den-cyberpeace/

Reinhold, T. (2013). Malware als Waffe. *ADLAS Magazin für Außen- und Sicherheitspolitik*, *13*(1), 7–11. https://adlasmagazin.files.wordpress.com/2013/02/adlas-0113.pdf

Reinhold, T. (2014b). Friedens- und Sicherheitspolitische Fragen zur Militarisierung des Cyberspace. *FIfF Kommunikation*, (4), 70–73. http://www.fiff.de/publikationen/fiff-kommunikation/fk-2014/fk-2014-4

Reinhold, T. (2014c). Internationale Kooperationsrichtlinien – ein Ausweg aus dem Attributionsdilemma. *Sicherheit und Frieden (S+F) / Security and Peace*, *32*(1), 23–27. https://www.jstor.org/stable/24233888

Reinhold, T. (2015a). Militarisierung des Cyberspace – Friedens- und Sicherheitspolitische Fragen. *Wissenschaft und Frieden*, (2), 31–34. https://wissenschaft-und-frieden.de/artikel/militarisierung-des-cyberspace/

Reinhold, T. (2015b). Möglichkeiten und Grenzen zur Bestimmung von Cyberwaffen. In Cunningham, Douglas, Hofstedt, Petra, Meer, Klaus, & Schmitt, Ingo (Eds.), *INFORMATIK 2015* (pp. 587–596, Vol. 246). Gesellschaft für Informatik e.V. http://dl.gi.de/handle/20.500.12116/2219

Reinhold, T. (2015c). Betrifft: Cyberpeace – Auswirkungen der Exportbeschränkungen von Cyberwaffen durch das Wassenaar-Abkommen. *FIfF Kommunikation*, (1), 6–7. https://www.fiff.de/publikationen/fiff-kommunikation/fk-2015/fk-2015-1/fk-2015-1-content/fk-1-15-s6.pdf

Reinhold, T. (2016a). Zur Verantwortung der Informatik in einer technologisierten Gesellschaft. *Sicherheit und Frieden*, *34*(4), 253–256. https://doi.org/10.5771/0175-274X-2016-4-253

Reinhold, T. (2016b). Der Cyberspace – Vorfälle, militärische Aufrüstung und erste Friedensbestrebungen. *Welttrends – das außenpolitische Journal*, *113*(3), 22–27.

Reinhold, T. (2016c). Die Bundeswehr zieht ins Cyberfeld – Ein Kommentar zum Aufbau des neuen Bundeswehr-Organisationsbereiches Cyber- und Informationsraum. *Blätter für deutsche und internationale Politik*, 17–20. https://www.blaetter.de/archiv/jahrgaenge/2016/juli/die-bundeswehr-zieht-ins-cyberfeld

Reinhold, T. (2016d). Cyberspace als Kriegsschauplatz? Herausforderungen für Völkerrecht und Sicherheitspolitik“ (APUZ 35–36/2016). *Aus Politik und Zeitgeschichte (APUZ), Bundeszentrale für politische Bildung*, *35–36*. http://www.bpb.de/apuz/232966/cyberspace-als-kriegsschauplatz

Reinhold, T. (2018b). Der Cyberspace ein neuer Kriegsschauplatz? Herausforderungen für Völkerrecht und Sicherheitspolitik. *Der Auftrag, Verbandszeitschrift der Gemeinschaft Katholischer Soldaten (GKS)*, *302*(2), 30–35. https://www.gemeinschaft-katholischer-soldaten.de/attachments/article/104/Auftrag_302_PDF_W.pdf

Reinhold, T. (2018c). Rethinking the Attribution Problem – A Plausible Proof of Non-Involvement as an Alternative to Credible Attribution [magazine]. *Issue brief 2: Briefing and memos from the research advisory group of the Global Commission on the Stability of Cyberspace (GCSC)*, (2), 134–149. https://hcss.nl/wp-content/uploads/2022/06/GCSC-Research-Advisory-Group-Issue-Brief-2-Bratislava.pdf

Reinhold, T. (2019a). Counting Cyber Weapons – New Approaches to Regulate and Control Destructive Cyber Tools. *SCIENCE PEACE SECURITY ’19 - Proceedings of the Interdisciplinary Conference on Technical Peace and Security Challenges*, 41–45. https://tuprints.ulb.tu-darmstadt.de/9164/2/2019_SciencePeaceSecurity_Proceedings-TUprints.pdf

Reinhold, T. (2019b). Rüstungskontrolle für den Cyberspace – Herausforderungen und erste Ansätze [magazine]. *FIfF-Kommunikation 3/2019 "Cyberpeace und IT-Security"*, (3), 26–29. https://www.fiff.de/publikationen/fiff-kommunikation/fk-jhrg-2019/FK-2019-3

Reinhold, T. (2019c, December 12). Cyberspace as Military Domain: Monitoring Cyberweapons. In D. Feldner (Ed.), *Redesigning Organizations: Concepts for the*

Connected Society (pp. 267–278). Springer International Publishing. https://doi.org/10.1007/978-3-030-27957-8_20

Reinhold, T. (2020a, June 9). *Russian Hacker Wanted!* Directions Blog. https://directionsblog.eu/russian-hacker-wanted/

Reinhold, T. (2020b, December 3). *Stellungnahme zur öffentlichen Anhörung im Verteidigungsausschuss des deutschen Bundestages zum Thema: Verfassungs- und völkerrechtliche Fragen im militärischen Cyber- und Informationsraum unter besonderer Berücksichtigung des Parlamentsvorbehalts, der Zurechenbarkeit von Cyberangriffen sowie einer möglichen Anpassung nationaler und internationaler Normen*. Deutscher Bundestag, Verteidigungsausschuss. https://www.bundestag.de/resource/blob/824622/67fc9db4f856a8445355562500d2a134/stellungnahme-Thomas-Reinhold_15-03-2021-data.pdf

Reinhold, T. (2021a). Der Weg zu einem sicheren zivilen Cyberspace. *Hoch3 - Die Zeitung der Technischen Universität Darmstadt*, *17*(10), 14. https://www.tu-darmstadt.de/universitaet/aktuelles_meldungen/publikationen/publikationen_archiv/einzelansicht_12672.de.jsp

Reinhold, T. (2021b). Überlegungen zur Militarisierung Künstlicher Intelligenz - Von Fallstricken, Grenzen und Problemen der Rüstungskontrolle. *Wissenschaft und Frieden - Dossier*, *4*(93). https://wissenschaft-und-frieden.de/dossier/kuenstliche-intelligenz-zieht-in-den-krieg/

Reinhold, T. (2021c). Zur Rolle und Verantwortung der Informatik für die Friedensforschung und Rüstungskontrolle. *FIfF Kommunikation*, (4), 47–49. https://www.fiff.de/publikationen/fiff-kommunikation/fk-2021/fk-2021-4/fk-2021-4-content/fk-4-21-p47.pdf

Reinhold, T. (2021d). Export Controls on Cyber-Surveillance Items. *Newsletter of the EU Non-Proliferation and Disarmament Consortium*, (33), 1. https://www.iai.it/sites/default/files/eunpd_e-newsletter_33.pdf

Reinhold, T. (2021e, December 30). *Export Control of Surveillance Software from Germany and Europe - Regulations, Limits and Weaknesses*. Heinrich-Böll-Stiftung. https://il.boell.org/en/2021/12/27/export-control-surveillance-software-germany-and-europe-regulations-limits-and

Reinhold, T. (2022). Arms Control for Artificial Intelligence. In T. Reinhold & N. Schörnig (Eds.), *Armament, Arms Control and Artificial Intelligence: The Janus-faced Nature of Machine Learning in the Military Realm* (pp. 211–226). Springer International Publishing. https://doi.org/10.1007/978-3-031-11043-6_15

Reinhold, T., Kühn, P., Günther, D., Schneider, T., & Reuter, C. (2023). EXTRUST: Reducing Exploit Stockpiles With a Privacy-Preserving Depletion System for Inter-State Relationship. *IEEE Transactions on Technology and Society*. https://doi.org/10.1109/TTS.2023.3280356

Reinhold, T., Pleil, H., & Reuter, C. (2023). Challenges for Cyber Arms Control: A Qualitative Expert Interview Study. *Zeitschrift für Außen- und Sicherheitspolitik (ZfAS)*. https://doi.org/10.1007/s12399-023-00960-w

Reinhold, T., & Reuter, C. (2019a, March 13). Arms Control and its Applicability to Cyberspace. In C. Reuter (Ed.), *Information Technology for Peace and Security: IT Applications and Infrastructures in Conflicts, Crises, War, and Peace* (pp. 207–231). Springer Vieweg. https://doi.org/10.1007/978-3-658-25652-4_10

Reinhold, T., & Reuter, C. (2019b, March 13). From Cyber War to Cyber Peace. In C. Reuter (Ed.), *Information Technology for Peace and Security: IT Applica-*

tions and Infrastructures in Conflicts, Crises, War, and Peace (pp. 139–164). Springer Vieweg. https://doi.org/10.1007/978-3-658-25652-4_7

Reinhold, T., & Reuter, C. (2019c, March 13). Verification in Cyberspace. In C. Reuter (Ed.), *Information Technology for Peace and Security: IT Applications and Infrastructures in Conflicts, Crises, War, and Peace* (pp. 257–275). Springer Vieweg. https://doi.org/10.1007/978-3-658-25652-4_12

Reinhold, T., & Reuter, C. (2021). Towards a Cyber Weapons Assessment Model – Assessment of the Technical Features of Malicious Software. *IEEE Transactions on Technology and Society*, *3*(3), 226–239. https://doi.org/10.1109/TTS.2021.3131817

Reinhold, T., & Reuter, C. (2022, October 9). Cyber Weapons and Artificial Intelligence: Impact, Influence and the Challenges for Arms Control. In T. Reinhold & N. Schörnig (Eds.), *Armament, Arms Control and Artificial Intelligence: The Janus-faced Nature of Machine Learning in the Military Realm* (pp. 145–158). Springer International Publishing. https://doi.org/10.1007/978-3-031-11043-6_11

Reinhold, T., & Reuter, C. (2023a). Preventing the Escalation of Cyber Conflicts: Towards an Approach To Plausibly Assure the Non-Involvement in a Cyberattack. *Zeitschrift für Friedens- und Konfliktforschung (ZeFKo)*. https://doi.org/10.1007/s42597-023-00099-7

Reinhold, T., & Reuter, C. (2023b). Zur Debatte über die Einhegung eines Cyberwars: Analyse militärischer Cyberaktivitäten im Krieg Russlands gegen die Ukraine. *Zeitschrift für Friedens- und Konfliktforschung*. https://doi.org/10.1007/s42597-023-00094-y

Reinhold, T., & Schörnig, N. (Eds.). (2022). *Armament, Arms Control and Artificial Intelligence: The Janus-faced Nature of Machine Learning in the Military Realm* (1st ed.). Springer Cham. https://doi.org/10.1007/978-3-031-11043-6

Reinhold, T., & Schulze, M. (2017). Digitale Gegenangriffe Eine Analyse der technischen und politischen Implikationen von „hack backs“. *SWP Arbeitspapier*, *01.08.2017*(1), 1–18. https://www.swp-berlin.org/publications/products/arbeitspapiere/AP_Schulze_Hackback_08_2017.pdf

Reuter, C. (Ed.). (2019). *Information Technology for Peace and Security: IT Applications and Infrastructures in Conflicts, Crises, War, and Peace*. Springer Fachmedien. https://doi.org/10.1007/978-3-658-25652-4

Reuter, C., Aldehoff, L., Riebe, T., & Kaufhold, M.-A. (2019). IT in Peace, Conflict, and Security Research. In C. Reuter (Ed.), *Information Technology for Peace and Security: IT Applications and Infrastructures in Conflicts, Crises, War, and Peace* (pp. 11–37). Springer Fachmedien. https://doi.org/10.1007/978-3-658-25652-4_2

Reuter, C., Altmann, J., Göttsche, M., & Himmel, M. (2020). Zur naturwissenschaftlich-technischen Friedens- und Konfliktforschung: Aktuelle Herausforderungen und Bewertung der Empfehlungen des Wissenschaftsrats. *Zeitschrift für Friedens- und Konfliktforschung*, *9*(1), 143–154. https://doi.org/10.1007/s42597-020-00035-z

Reuter, C., Riebe, T., Aldehoff, L., Kaufhold, M.-A., & Reinhold, T. (2019). Cyberwar zwischen Fiktion und Realität – technologische Möglichkeiten. In I.-J. Werkner & N. Schörnig (Eds.), *Cyberwar – die Digitalisierung der Kriegsführung: Fragen zur Gewalt. Band 6* (pp. 15–38). Springer VS. https://doi.org/10.1007/978-3-658-27713-0_2

Reuter, C., Riebe, T., Haunschild, J., Reinhold, T., & Schmid, S. (2022). Zur Schnittmenge von Informatik mit Friedens- und Sicherheitsforschung: Erfahrungen aus der interdisziplinären Lehre in der Friedensinformatik. *Zeitschrift für Friedens- und Konfliktforschung*, *11*(2), 129–140. https://doi.org/10.1007/s42597-022-00078-4

Rid, T., & Buchanan, B. (2015). Attributing Cyber Attacks. *Journal of Strategic Studies*, *38*, 4–37. https://doi.org/10.1080/01402390.2014.977382

Rid, T., & McBurney, P. (2012). Cyber-Weapons. *The RUSI Journal*, *157*(1), 6–13. https://doi.org/10.1080/03071847.2012.664354

Riebe, T., Kaufhold, M.-A., Kumar, T., & Reuter, C. (2019). Threat Intelligence Application for Cyber Attribution. In C. Reuter (Ed.), *SCIENCE PEACE SECURITY '19 - Proceedings of the Interdisciplinary Conference on Technical Peace and Security Challenges* (pp. 56–59). TUprints. https://tubiblio.ulb.tu-darmstadt.de/116000/

Riebe, T., & Reuter, C. (2019a). Dual-Use and Dilemmas for Cybersecurity, Peace and Technology Assessment. In C. Reuter (Ed.), *Information Technology for Peace and Security: IT Applications and Infrastructures in Conflicts, Crises, War, and Peace* (pp. 165–183). Springer Fachmedien. https://doi.org/10.1007/978-3-658-25652-4_8

Riebe, T., & Reuter, C. (2019b). Dual-Use in der IT. *Wissenschaft & Frieden*, *01*. https://wissenschaft-und-frieden.de/artikel/dual-use-in-der-it/

Roberts, P. (2019, December 13). *AI for Peace*. War on the Rocks. https://warontherocks.com/2019/12/ai-for-peace/

Robertson, L. J., Munoz, A., & Michael, K. (2020). Managing Technological Vulnerability of Urban Dwellers: Analysis, Trends, and Solutions. *IEEE Transactions on Technology and Society*, *1*(1), 48–59. https://doi.org/10.1109/TTS.2020.2975806

Rõigas, H., & Minárik, T. (2015). 2015 UN GGE Report: Major Players Recommending Norms of Behaviour, Highlighting Aspects of International Law. *NATO Cooperative Cyber Defence Centre of Excellence, August*, *31*. https://ccdcoe.org/incyder-articles/2015-un-gge-report-major-players-recommending-norms-of-behaviour-highlighting-aspects-of-international-law/

Roßner, S. (2017). Die Ächtung von Waffen: Abkommen der qualitativen Rüstungsbegrenzung. In I.-J. Werkner & K. Ebeling (Eds.), *Handbuch Friedensethik* (pp. 769–780). Springer Fachmedien. https://doi.org/10.1007/978-3-658-14686-3_55

Rovner, J. (2020). *The Intelligence Contest in Cyberspace*. Lawfare. https://www.lawfareblog.com/intelligence-contest-cyberspace

Rovnyagin, M. M., Dmitriev, S. O., Hrapov, A. S., Maksutov, A. A., & Turovskiy, I. A. (2021). Database Storage Format for High Performance Analytics of Immutable Data. *2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus)*, 618–622. https://doi.org/10.1109/ElConRus51938.2021.9396453

Rowe, N. (2015). The attribution of cyber warfare. In J. A. Green (Ed.), *Cyber warfare: A multidisciplinary analysis* (pp. 61–72). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315761565-4/attribution-cyber-warfare-1-neil-rowe

Rubio, J. E., Alcaraz, C., Roman, R., & Lopez, J. (2019). Current cyber-defense trends in industrial control systems. *Computers & Security*, *87*(101561), 101561. https://doi.org/10.1016/j.cose.2019.06.015

Rueda, D. F., Calle, E., & Marzo, J. L. (2017). Robustness Comparison of 15 Real Telecommunication Networks: Structural and Centrality Measurements. *Journal of Network and Systems Management*, *25*(2), 269–289. https://doi.org/10.1007/s10922-016-9391-y

Rüger, C. (2012). *Europäische Außen- und Sicherheitspolitik – (k)ein Thema für die Öffentlichkeit? Die außen- und sicherheitspolitische Rolle der EU im Blickwinkel der öffentlichen Meinung und Medien*. Nomos Verlag. https://www.nomos-shop.de/nomos/titel/europaeische-aussen-und-sicherheitspolitik-kein-thema-fuer-die-oeffentlichkeit-id-79349/

Ruhmann, I. (2010). Rüstungskontrolle gegen den Cyberkrieg? http://www.heise.de/tp/druck/ob/artikel/31/31797/1.html

Ruhmann, I. (2015). *Neue Ansätze für die Rüstungskontrolle bei Cyber-Konflikten*. Gesellschaft für Informatik e.V. http://dl.gi.de/handle/20.500.12116/2218

Rupprecht, D., Dabrowski, A., Holz, T., Weippl, E., & Popper, C. (2018). On security research towards future mobile network generations. *IEEE Communications Surveys and Tutorials*, *20*(3), 2518–2542. https://doi.org/10.1109/COMST.2018.2820728

Russell, B. (2020). IoT Cyber Security. In F. Firouzi, K. Chakrabarty, & S. Nassif (Eds.), *Intelligent Internet of Things: From Device to Fog and Cloud* (pp. 473–512). Springer International Publishing. https://doi.org/10.1007/978-3-030-30367-9_10

Russell, R. L., Kim, L., Hamilton, L. H., Lazovich, T., Harer, J. A., Ozdemir, O., Ellingwood, P. M., & McConley, M. W. (2018, November 28). *Automated Vulnerability Detection in Source Code Using Deep Representation Learning*. https://doi.org/10.48550/arXiv.1807.04320

Sadique, F., Cheung, S., Vakilinia, I., Badsha, S., & Sengupta, S. (2018). Automated structured threat information expression (STIX) document generation with privacy preservation. *Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON'18)*, 847–853. https://doi.org/10.1109/UEMCON.2018.8796822

Sander, B. (2019). The Sound of Silence: International Law and the Governance of Peacetime Cyber Operations. *2019 11th International Conference on Cyber Conflict (CyCon)*, *900*, 1–21. https://doi.org/10.23919/CYCON.2019.8756882

Sandholz, S., Sett, D., & Saenge, N. (2020). Critical infrastructure failures: How resilient is the population? *Population Protection*, *1*, 32–35. https://www.bbk.bund.de/SharedDocs/Downloads/DE/Mediathek/Publikationen/BSMAG/bsmag_20_1.pdf?__blob=publicationFile&v=4.

Sanger, D. E. (2014). Syria War Stirs New U.S. Debate on Cyberattacks [newspaper]. *The New York Times: World*. https://www.nytimes.com/2014/02/25/world/middleeast/obama-worried-about-effects-of-waging-cyberwar-in-syria.html

Sanger, D. E. (2016). Obama Strikes Back at Russia for Election Hacking [newspaper]. *The New York Times: U.S.* https://www.nytimes.com/2016/12/29/us/politics/russia-election-hacking-sanctions.html

Sauerwein, C., Sillaber, C., Mussmann, A., & Breu, R. (2017). Threat intelligence sharing platforms: An exploratory study of software vendors and research perspectives. *Proceedings Der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017)*, 837–851. https://www.wi2017.ch/images/wi2017-0188.pdf

Scaglia, S. (2007). *The Embedded Internet: TCP/IP Basics, Implementation and Applications*. Addison-Wesley Professional. https://dl.acm.org/citation.cfm?id=1512809.

Schlangenstein, M. (2017). FedEx Cuts Profit Forecast on $300 Million Hit From Cyberattack [newspaper]. *Bloomberg.com*. https://www.bloomberg.com/news/articles/2017-09-19/fedex-cuts-profit-outlook-on-300-million-blow-from-cyberattack

Schneier, B. (1996). *Applied cryptography - protocols, algorithms, and source code in C*. John Wiley & Sons. https://www.schneier.com/books/applied-cryptography/

Schneier, B., & Kelsey, J. (1998). Cryptographic support for secure logs on untrusted machines. *Proceedings of the 7th Conference on USENIX Security Symposium - Volume 7*, *7*, 4. http://dl.acm.org/citation.cfm?id=1267549.1267553

Schörnig, N. (2018). Artificial intelligence in the military: More than killer robots. In B. Wolff (Ed.), *Whither artificial intelligence ? Debating the policy challenges of the upcoming transformation* (3rd, pp. 39–44). Science Policy Paper des Mercator Science-Policy Fellowship-Programms. https://www.stiftung-mercator.de/content/uploads/2020/12/PolicyPaper3_Artificial_Intelligence.pdf#page=42

Schörnig, N., & Reinhold, T. (2022). Introduction. In T. Reinhold & N. Schörnig (Eds.), *Armament, Arms Control and Artificial Intelligence: The Janus-faced Nature of Machine Learning in the Military Realm* (pp. 1–9). Springer International Publishing. https://doi.org/10.1007/978-3-031-11043-6_1

Schulze, M., & Reinhold, T. (2018). Wannacry About the Tragedy of the Commons? Game-Theory and the Failure of Global Vulnerability Disclosure. *European Conference on Information Warfare and Security, ECCWS*, *2018-June*, 454–463. https://www.proquest.com/openview/f6ccddd62973bd8997c3fcd40951f4f1/1?cbl=396497&pq-origsite=gscholar&parentSessionId=7jm9tc94UMKaTk9pAtTjzd%2BhJYdl8V55qGHqrUpnUM8%3D

Schulze, M. (2019). Quo Vadis Cyber Arms Control? – A Sketch of an International Vulnerability Equities Process and a 0-Day Emissions Trading Regime. *Proceedings of the Interdisciplinary Conference on Technical Peace and Security Research 2019*.

Schulze, M. (2022). Cyber operations in the context of the 2022 Russia-Ukraine war. *Ukraine Analyses*, *267*, 2–7. https://www.laender-analysen.de/ukraine-analysen/267/UkraineAnalysen267.pdf.

Schulze, M., & Datzer, V. (2021). *Neuer UN GGE Report zu Cyber-Fragen: Vorwärts mit kleinen Schritten*. Stiftung Wissenschaft und Politik (SWP). https://www.swp-berlin.org/publikation/neuer-un-gge-report-zu-cyber-fragen-vorwaerts-mit-kleinen-schritten

Schwarz, E. (2019). Günther Anders in Silicon Valley: Artificial intelligence and moral atrophy. *Thesis Eleven*, *153*(1), 94–112. https://doi.org/10.1177/0725513619863854

Schwirtz, M., Troianovski, A., Al-Hlou, Y., Froliak, M., Entous, A., & Gibbons-Neff, T. (2022). Putin's War: The Inside Story of a Catastrophe [newspaper]. *The New York Times: World*. https://www.nytimes.com/interactive/2022/12/16/world/europe/russia-putin-war-failures-ukraine.html

Seals, T. (2022). Relentless russian cyberattacks on ukraine raise important policy questions. *Dark Reading October*, *5*. https://www.darkreading.com/threat-intelligence/russian-cyberattacks-ukraine-raise-important-policy-questions

Security, O. for, & Europe, C.-o. in. (2016). Decision no. 1202. OSCE confidence-building measures to reduce the risks of conflict stemming from the use of information and communication technologies. https://www.osce.org/files/f/documents/d/a/227281.pdf.

Segal, A. (2017). *The development of cyber norms at the united nations ends in deadlock. Now what?* https://www.cfr.org/blog/development-cyber-norms-united-nations-ends-deadlock-now-what

Shacham, H. (2007). The geometry of innocent flesh on the bone: Return-into-libc without function calls (on the x86). *Proceedings of the 14th ACM Conference on Computer and Communications Security*, 552–561. https://doi.org/10.1145/1315245.1315313

Shachtman, N. (2009). Kremlin Kids: We Launched the Estonian Cyber War [magazine]. *Wired*. https://www.wired.com/2009/03/pro-kremlin-gro/

Shea, J. (2018). Cyberspace as a domain of operations: What is nato's vision and strategy? *MCU Journal*, *9*(2). https://doi.org/10.21140/mcuj.2018090208

Shepherd, S. (2003). Vulnerability disclosure. how do we define responsible disclosure? https://www.sans.org/white-papers/932/

Sherry, L. (1996). Supporting a networked community of learners. *TechTrends : for leaders in education & training*, *41*(4), 28–32. https://link.springer.com/content/pdf/10.1007/BF02818903.pdf

Shute, S., Ko, R. K. L., & Chaisiri, S. (2017). Attribution Using Keyboard Row Based Behavioural Biometrics for Handedness Recognition. *2017 IEEE Trustcom/BigDataSE/ICESS*, 1131–1138. https://doi.org/10.1109/Trustcom/BigDataSE/ICESS.2017.363

Sibi Chakkaravarthy, S., Sangeetha, D., & Vaidehi, V. (2019). A Survey on malware analysis and mitigation techniques. *Computer Science Review*, *32*, 1–23. https://doi.org/10.1016/j.cosrev.2019.01.002

Sigholm, J. (2013). Non-State Actors in Cyberspace Operations. *Journal of Military Studies*, *4*(1), 1–37. https://doi.org/10.1515/jms-2016-0184

Silomon, J. (2018). Software as a Weapon: Factors Contributing to the Development and Proliferation. *Journal of Information Warfare*, *17*(3), 106–123. https://www.jstor.org/stable/26633169

Silomon, J. A. (2020). Bug bounties: Bottom-up initiatives as forms of cyber arms control? *Proceedings of the 15th International Conference on Cyber Warfare and Security, ICCWS 2020*, 431–438. https://doi.org/10.34190/ICCWS.20.137

Singer, P., & Friedman, A. (2014, January 15). *Cult of the Cyber Offensive*. Foreign Policy. https://foreignpolicy.com/2014/01/15/cult-of-the-cyber-offensive/

Sir Amyas Morse KCB. (2018). *Investigation: WannaCry cyber attack and the NH*. UK National Audit Office. https://www.nao.org.uk/wp-content/uploads/2017/10/Investigation-WannaCry-cyber-attack-and-the-NHS.pdf

Smith, B. (2017, February 14). *The need for a digital geneva convention*. https://blogs.microsoft.com/on-the-issues/2017/02/14/need-digital-geneva-convention/

Sommer, P., & Brown, I. (2011, January 14). *Reducing Systemic Cybersecurity Risk*. https://papers.ssrn.com/abstract=1743384

South, A. (2011). Rworldmap : A new R package for mapping global data. *The R Journal*, *3*(1), 35–43. https://doi.org/10.32614/RJ-2011-006

Spiegel Staff. (2018). The Breach from the East [newspaper]. *Der Spiegel Online: International*. https://www.spiegel.de/international/germany/cyber-espionage-likely-from-russia-targets-german-government-a-1196520.html

Staff, B. (2020, December 30). Several institutions affected by email virus in Lithuania – center. https://www.baltictimes.com/several_institutions_affected_by_email_virus_in_lithuania___center/

Starski, P. D. P. (2015, September 18). *Right to Self-Defence, Attribution and the Non-State Actor – Birth of the 'Unable and Unwilling' Standard?* https://doi.org/10.2139/ssrn.2692422

Steffens, T. (2020). *Attribution of Advanced Persistent Threats: How to Identify the Actors Behind Cyber-Espionage*. Springer. https://doi.org/10.1007/978-3-662-61313-9

Stein, A. A. (1982). Coordination and Collaboration: Regimes in an Anarchic World. *International Organization*, *36*(2), 299–324. https://www.jstor.org/stable/2706524

Stergiopoulos, G., Kotzanikolaou, P., Theocharidou, M., & Gritzalis, D. (2015). Risk mitigation strategies for critical infrastructures based on graph centrality analysis. *International Journal of Critical Infrastructure Protection*, *10*, 34–44. https://doi.org/10.1016/j.ijcip.2015.05.003

Stevens, T. (2018). Cyberweapons: Power and the governance of the invisible. *International Politics*, *55*(3), 482–502. https://doi.org/10.1057/s41311-017-0088-y

Stohl, R., & Grillot, S. (2009). *The International Arms Trade*. Polity Press. https://www.wiley.com/en-ie/The+International+Arms+Trade-p-9780745641546

SubTelForum. (2020a). Submarine cable almanac. https://subtelforum.com/products/submarine-cable-almanac/

SubTelForum. (2020b). Submarine telecoms industry report 2020/2021 edition. https://subtelforum.com/products/submarine-telecoms-industry-report/

Sutherland, E. (2009, August 15). *Telecommunications in Small Island Developing States*. https://doi.org/10.2139/ssrn.1469243

Symantec. (2017). *Internet Security Threat Report. Ransomware 2017*. https://docs.broadcom.com/doc/istr-22-2017-en

Tang, S. (2009). The Security Dilemma: A Conceptual Analysis. *Security Studies*, *18*(3), 587–623. https://doi.org/10.1080/09636410903133050

Tanriverdi, H. (2018). This is how the hackers smuggled data out of the German Foreign Office [newspaper]. *Süddeutsche Zeitung*. https://www.sueddeutsche.de/digital/exklusiv-so-schleusten-die-hacker-daten-aus-dem-auswaertigen-amt-1.3894534

Tarakanov, Dimitry. (2022). *Shamoon The Wiper: Further Details (Part II)*. SecureList. https://securelist.com/shamoon-the-wiper-further-details-part-ii/57784/

TeleGeography. (2020). Submarine cable map. https://www.submarinecablemap.com/#/

TeleGeography. (2021). The state of the network. https://www2.telegeography.com/hubfs/assets/Ebooks/state-of-the-network-2021.pdf

The Editorial Board. (2015). Opinion | Arms Control for a Cyberage [newspaper]. *The New York Times: Opinion*. https://www.nytimes.com/2015/02/26/opinion/arms-control-for-a-cyberage.html

The Hague Conference. (1899). Convention (II) with respect to the laws and customs of war on land and its annex: Regulations concerning the laws and customs of war on land. *The Hague*. https://ihl-databases.icrc.org/assets/treaties/150-IHL-10-EN.pdf

The Hague Conference. (1907). Convention (IV) respecting the Laws and Customs of War on Land and its annex: Regulations concerning the Laws and Customs of War on Land. https://ihl-databases.icrc.org/ihl/INTRO/195

The State Council Information Office of the People's Republic of China. (2019). *The State Council Information Office of the People's Republic of China* (Report.). Foreign Languages Press. Beijing. http://www.xinhuanet.com/english/download/whitepaperonnationaldefenseinnewera.doc

The Tor Project. (2019). *Tor Project: Overview*. https://2019.www.torproject.org/about/overview.html.en#overview.

The Wassenaar Arrangement Secretariat. (2014). Recommendations for the Implementation of the 2013 Wassenaar Arrangement Changes Regarding "Intrusion Software" and "IP Network Communications Surveillance Systems". https://static.newamerica.org/attachments/541-human-rights-and-technology-organizations-submit-joint-recommendations-to-the-us-government-on-the-implementation-of-the-2013-wassenaar-amendments-on-surveillance-technology/Joint_Recommendations_Wassenaar_Implementation.c6e0fc99218b4384b41d705038c8f8eb.pdf

The Wassenaar Arrangement Secretariat. (2019). Wassenaar Arrangement on Export Controls for Conventional Arms and Dual-use Goods and Technologies - Founding Documents. https://www.wassenaar.org/app/uploads/2021/12/Public-Docs-Vol-I-Founding-Documents.pdf

The Wassenaar Arrangement Secretariat. (2022). Wassenaar Arrangement on Export Controls for Conventional Arms and Dual-use Goods and Technologies - List of Dual-Use Goods and Technologies and Munitions list (Version December 2022). https://www.wassenaar.org/app/uploads/2022/12/List-of-Dual-Use-Goods-and-Technologies-Munitions-List-Dec-2022.pdf

Thompson, J. R., Frezza, D., Necioglu, B., Cohen, M. L., Hoffman, K., & Rosfjord, K. (2019). Interdependent Critical Infrastructure Model (ICIM): An agent-based model of power and water infrastructure. *International Journal of Critical Infrastructure Protection*, *24*, 144–165. https://doi.org/10.1016/j.ijcip.2018.12.002

Thorat, D. (2019). Colonial Topographies of Internet Infrastructure: The Sedimented and Linked Networks of the Telegraph and Submarine Fiber Optic Internet. *South Asian Review*, *40*(3), 252–267. https://doi.org/10.1080/02759527.2019.1599563

Threatpost. (2017). EternalBlue exploit used in retefe banking trojan campaign. https://threatpost.com/eternalblue-exploit-used-in-retefe-banking-trojan-campaign/128103/.

Tidy, J. (2022). Ukraine: EU deploys cyber rapid-response team [newspaper]. *BBC News: Technology*. https://www.bbc.com/news/technology-60484979

Tikk, E., & Kerttunen, M. (2017). The alleged demise of the UN GGE: An autopsy and eulogy. http://cpi.ee/wp-content/uploads/2017/12/2017-Tikk-Kerttunen-Demise-of-the-UN-GGE-2017-12-17-ET.pdf

Tikk-Ringas, E. (2012). Developments in the field of information and telecommunication in the con-text of international security: Work of the UN first Committee 1998—2012. https://citizenlab.ca/cybernorms2012/ungge.pdf

Trendmicro. (2017). Midyear security roundup: The cost of compromise - security roundup - trend micro GB. https://www.trendmicro.com/vinfo/us/security/research-and-analysis/threat-reports/roundup/the-cost-of-compromise

Trevithick, J. (2019). Ukrainian Officer Details Russian Electronic Warfare Tactics Including Radio "Virus". *The Drive. The WarZone. October*, 1–13. https://www.thedrive.com/the-war-zone/30741/ukrainian-officer-details-russian-electronic-warfare-tactics-including-radio-virus.

Tucker, J. B. (1998). Verification Provisions of the Chemical Weapons Convention and Their Relevance to the Biological Weapons Convention - Biological Weapons Proliferation: Reasons for Concern, Courses of Action. *Stimson Center Report*, *24*. http://www.acamedia.info/politics/IRef/StimsonC/report24-tucker.PDF

UN. (1963). Treaty banning nuclear weapon tests in the atmosphere, in outer space and under water (partial test ban treaty. https://treaties.un.org/doc/Publication/UNTS/Volume%20480/volume-480-I-6964-English.pdf

UN. (1967). Treaty on principles governing the activities of states in the exploration and use of outer space, including the moon and other celestial bodies. http://www.unoosa.org/pdf/publications/STSPACE11E.pdf

UN. (1969). Vienna convention on the law of treaties. United nations. https://treaties.un.org/doc/publication/unts/volume%201155/volume-1155-i-18232-english.pdf

UN. (1972a). Convention on the prohibition of the development, production and stockpiling of bacteriological (biological) and toxin weapons and on their destruction (btwc. https://legal.un.org/avl/ha/cpdpsbbtwd/cpdpsbbtwd.html

UN. (1972b). Treaty between the united states of america and union of soviet socialist republics on the limitation of anti-ballistic missile systems. https://treaties.un.org/doc/Publication/UNTS/Volume%20944/volume-944-I-13446-English.pdf

UN. (1988). General Assembly, Special Report of the Disarmament Commission to the General Assembly at its Third Special Session Devoted to Disarmament, UN document A/S-15/3, 28 May 1988. http://undocs.org/en/A/S-15/3(SUPP)

UN. (1999). A/Res/53/70 developments in the field of information and telecommunications in the context of international security. https://digitallibrary.un.org/record/265311

UN. (2003). Resolution adopted by the General Assembly on 8 December 2003 on Developments in the field of information and telecommunications in the context of international security. https://ccdcoe.org/uploads/2018/10/UN-031208-ITIS_0.pdf

UN. (2013). Arms trade treaty. https://treaties.un.org/Pages/ViewDetails.aspx?src=IND&mtdsg_no=XXVI-8&chapter=26&clang=_en

UN. (2016). Report of the open-ended intergovernmental expert working group on indicators and terminology relating to disaster risk reduction. https://undocs.org/en/A/71/644

UN. (2018a). Advancing responsible state behaviour in cyberspace in the context of international security (A/C.1/73/L.37. http://undocs.org/A/C.1/73/L.37

UN. (2018b). Draft resolution by russia and other states concerning the developments in the field of information and telecommunications in the context of international security. http://undocs.org/A/C.1/73/L.27

UN General Assembly. (2011). Letter dated 12 september 2011 from the permanent representatives of chin, the russian federation, tajikistan and uzbekistan to the united nations addressed to the secretary-general. https://digitallibrary.un.org/record/710973

UN-ESCAP. (2018). Broadband connectivity in pacific island countries. https://repository.unescap.org/bitstream/handle/20.500.12870/1315/ESCAP-2018-WP-Broadband-Connectivity-Pacific.pdf?sequence=1&isAllowed=y

UNGGE. (2015a). Consensus report 2015 - Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security - A/70/174. http://undocs.org/A/70/174

UNGGE. (2015b). Report of the group of governmental experts on developments in the field of information and telecommunications in the context of international security. https://digitallibrary.un.org/record/799853

UNGGE. (2019). Draft Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons

Systems. https://documents.unoda.org/wp-content/uploads/2020/09/CCW_GGE.1_2019_3_E.pdf

UNGGE. (2021). *Report of the group of governmental experts on advancing responsible state behaviour in cyberspace in the context of international security to the united nations general assembly*. https://front.un-arm.org/wp-content/uploads/2021/08/A_76_135-2104030E-1.pdf

UNHCR. (2019). *Human rights in the digital age-Can they make a difference?* https://www.ohchr.org/en/speeches/2019/10/human-rights-digital-age

UNIDIR. (2013). The Cyber Index - International Security Trends and Realities. https://unidir.org/sites/default/files/publication/pdfs//cyber-index-2013-en-463.pdf

UNIDIR. (2016). Report of the international security cyber issues workshop series. https://unidir.org/publication/report-international-security-cyber-issues-workshop-series

UNIDIR. (2018). Preventing and mitigating ICT-Related conflict. *Cyber Stability Conference 2018 Summary Report*. https://www.unidir.org/publication/preventing-and-mitigating-ict-related-conflict-cyber-stability-conference-2018-summary.

Union of Concerned Scientists. (2022, May 1). UCS satellite database. https://www.ucsusa.org/resources/satellite-database

UNODA. (2017). Voluntary, non-binding norms for responsible state behaviour in the use of information and communications technology: A commentary. https://www.un.org/disarmament/wp-content/uploads/2018/04/Civil-Society-2017.pdf

UNODA. (2021). Final Session of the Group of Governmental Experts on Advancing responsible State behavior in cyberspace in the context of international security - Remarks by Ms. Izumi Nakamitsu High Representative for Disarmament Affairs. *United Nations Office for Disarmament Affairs. May*, *28*. https://front.un-arm.org/wp-content/uploads/2021/05/HR-remarks-at-Final-Session-of-the-Group-of-Governmental-Experts-on-Advancing-responsible-State-behaviour-in-cyberspace-in-the-context-of-international-security.pdf

UNODA. (2022). Open-ended working group on security of and in the use of information and communications technologies 2021–2025, Third substantive session New York, 25–29 July 2022 Provisional programme of work. https://s3-eu-west-1.amazonaws.com/upload.teamup.com/908040/TuosiLDlTjWVPdmXEFBY_PoW_3rd-20Session_ODS.pdf

US Air Force Secretary. (2018). Air Force Instruction 51-401. In *The Law of War*. https://static.e-publishing.af.mil/production/1/af_ja/publication/afi51-401/afi51-401.pdf

US Department of Justice. (2018). Case 1:18-cr-00032-DLF. *Case*, *1*(18). https://www.justice.gov/file/1035477/download

US Department of Justice. (2020). Six russian GRU officers charged in connection with worldwide deployment of destructive malware and other disruptive actions in cyberspace. https://www.justice.gov/opa/pr/six-russian-gru-officers-charged-connection-worldwide-deployment-destructive-malware-and

US White House. (2016, April 13). Statement by the president on progress in the fight against ISIL. www.whitehouse.gov/the-press-office/2016/04/13/statement-president-progress-fight-against-isil

US White House. (2017). Vulnerabilities Equities Policy and Process for the United States Government. https://trumpwhitehouse.archives.gov/sites/whitehouse.gov/files/images/External%20-%20Unclassified%20VEP%20Charter%20FINAL.PDF

US White House. (2021). Remarks by president biden at the office of the director of national intelligence. https://www.whitehouse.gov/briefing-room/speeches-remarks/2021/07/27/remarks-by-president-biden-at-the-office-of-the-director-of-national-intelligence/.

US-DHS. (2020). Guidance on the north korean cyber threat. https://www.us-cert.gov/ncas/alerts/aa20-106a

US-DNI. (2017, January 6). Background to "Assessing Russian Activities and Intentions in Recent US Elections": The analytic process and cyber incident attribution. https://www.dni.gov/files/documents/ICA_2017_01.pdf

US-DOD. (2018a). Achieve and Maintain Cyberspace Superiority Command Vision for US Cyber Command, 1–12. https://www.cybercom.mil/Portals/56/Documents/USCYBERCOM%20Vision%20April%202018.pdf

US-DOD. (2018b). National cyber strategy. https://trumpwhitehouse.archives.gov/wp-content/uploads/2018/09/National-Cyber-Strategy.pdf

US-DOD. (2018c). Summary of the 2018 department of defense AI strategy. https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF

Van Dijk, J. (2017). Afterword: The state of digital divide theory. In *Theorizing digital divides* (pp. 199–206). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315455334-16/afterword-jan-van-dijk

Van Eeten, M., Nieuwenhuijs, A., Luiijf, E., Klaver, M., & Cruz, E. (2011). The state and the threat of cascading failure across critical infrastructures: The implications of empirical evidence from media incident reports. *Public Administration*, *89*(2), 381–400. https://doi.org/10.1111/j.1467-9299.2011.01926.x

VERIS. (2019). Vocabulary for Event Recording and Incident Sharing Framework. http://veriscommunity.net/

Verizon. (2018). *Data breach investigations report*. https://www.researchgate.net/publication/324455350_2018_Verizon_Data_Breach_Investigations_Report

Verizon. (2019). 2019 Data Breach Investigations Report. *Verizon Business Journal*. https://doi.org/10.1016/S1361-3723(19)30060-0

Voell, D., Gaski, F. L.-N., Jagadeesan, R., Khasanshyn, R., Montgomery, H., Teis, S., Blummer, T., Katipalli, M. K., & Bowman, M. (2016). Hyperledger whitepaper. https://www.hyperledger.org/wp-content/uploads/2018/08/HL_Whitepaper_IntroductiontoHyperledger.pdf

Voo, J., Hemani, I., Jones, S., DeSombre, W., Cassidy, D., & Schwarzenbach, A. (2020). *National cyber power index 2020*. Harvard Kennedy School, Belfer Center for Science and International Affairs. Cambridge, MA. https://www.belfercenter.org/sites/default/files/2020-09/NCPI_2020.pdf

Wagner, T. D., Mahbub, K., Palomar, E., & Abdallah, A. E. (2019). Cyber threat intelligence sharing: Survey and research directions. *Computers & Security*, *87*, 101589. https://doi.org/10.1016/j.cose.2019.101589

Waksman, A. (1968). A Permutation Network. *Journal of the ACM*, *15*(1), 159–163. https://doi.org/10.1145/321439.321449

Warrell, H., & Foy, H. (2019). Russian cyberattack unit ‘masqueraded’ as iranian hackers, uk says. *Financial Times*. https://www.ft.com/content/b947b46a-f342-11e9-a79c-bc9acae3b654.

Wehberg, H. (1959). Pacta Sunt Servanda. *American Journal of International Law*, *53*(4), 775–786. https://doi.org/10.2307/2195750

Weinberg, A. (2021). Analysis of top 11 cyber attacks on critical infrastructure. *First Point Blog*. https://www.firstpoint-mg.com/blog/analysis-of-top-11-cyber-attackson-critical-infrastructure/

Werkner, I.-J., & Schörnig, N. (Eds.). (2019). *Cyberwar – die Digitalisierung der Kriegsführung: Fragen zur Gewalt • Band 6*. Springer Fachmedien. https://doi.org/10.1007/978-3-658-27713-0

Wikipedia. (2023, March 8). List of cyber warfare forces. In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=List_of_cyber_warfare_forces&oldid=1143506610

Wingfield, T., & Wingo, H. (2021, November 4). International Law for Cyberspace: Competition and Conflict. In P. Cornish (Ed.), *The Oxford Handbook of Cyber Security* (pp. 578–594). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198800682.013.37

Winseck, D. (2017). The Geopolitical Economy of the Global Internet Infrastructure. *Journal of Information Policy*, *7*, 228–267. https://doi.org/10.5325/jinfopoli.7.2017.0228

Wisotzki, S., & Mutschler, M. (2021). No common position! European arms export control in crisis. *Zeitschrift für Friedens- und Konfliktforschung*, *10*(2), 273–293. https://doi.org/10.1007/s42597-022-00071-x

Wissenschaftsrat Deutschlands. (2019). *Empfehlungen zur Weiterentwicklung der Friedens- und Konfliktforschung* (Report,). Drs. Gießen. https://www.wissenschaftsrat.de/download/2019/7827-19.pdf?__blob=publicationFile&v=7

Woolf, A. (2010). *Monitoring and verification in arms control* (CRS report for congress.). https://www.nti.org/media/pdfs/Monitoring_and_Verification_in_Arms_Control.pdf

Wrozek, B. (2017). *Cyber Kill Chain Methodology*. https://www.scribd.com/document/410345440/Cyber-Kill-Chain-Wrozek-pdf

Xie, A., Wang, X., & Lu, S. (2019). Risk minimization routing against geographically correlated failures. *IEEE access : practical innovations, open solutions*, *7*, 62920–62929. https://doi.org/10.1109/ACCESS.2019.2916834

Yao, A. C.-C. (1986). How to generate and exchange secrets. *27th Annual Symposium on Foundations of Computer Science (Sfcs 1986)*, 162–167. https://doi.org/10.1109/SFCS.1986.25

Ye, C., Li, G., Cai, H., Gu, Y., & Fukuda, A. (2018). Analysis of Security in Blockchain: Case Study in 51%-Attack Detecting. *2018 5th International Conference on Dependable Systems and Their Applications (DSA)*, 15–24. https://doi.org/10.1109/DSA.2018.00015

You, W., Zong, P., Chen, K., Wang, X., Liao, X., Bian, P., & Liang, B. (2017). SemFuzz: Semantics-based Automatic Generation of Proof-of-Concept Exploits. *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2139–2154. https://doi.org/10.1145/3133956.3134085

Young, O. R. (1980). International Regimes: Problems of Concept Formation. *World Politics*, *32*(3), 331–356. https://doi.org/10.2307/2010108

Young, O. R. (1982). Regime dynamics: The rise and fall of international regimes. *International Organization*, *36*(2), 277–297. https://doi.org/10.1017/S0020818300018956

Zangl, B., & Zürn, M. (1994). Theorien des rationalen Handelns in den Internationalen Beziehungen. In V. Kunz & U. Druwe (Eds.), *Rational Choice in der Politikwissenschaft: Grundlagen und Anwendungen* (pp. 81–111). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-663-01128-6_5

Zeadally, S., & Flowers, A. (2014). Cyberwar: The What, When, Why, and How [Commentary]. *IEEE Technology and Society Magazine*, *33*(3), 14–21. https://doi.org/10.1109/MTS.2014.2345196

ZERODIUM. (2019). Zerodium - The leading exploit acquisition platform for premium zero-days and advanced cybersecurity capabilities. https://www.zerodium.com/

Zetter, K. (2014). Meet MonsterMind, the NSA Bot That Could Wage Cyberwar Autonomously [magazine]. *Wired*. https://www.wired.com/2014/08/nsa-monstermind-cyberwarfare/

Zetter, K. (2016). Inside the Cunning, Unprecedented Hack of Ukraine's Power Grid [magazine]. *Wired*. https://www.wired.com/2016/03/inside-cunning-unprecedented-hack-ukraines-power-grid/

Zetter, K. (2018). Leaked files show how the NSA tracks other countries' hackers. *The Intercept*. https://theintercept.com/2018/03/06/leaked-files-show-how-nsa-tracks-other-countries-hackers/

Zhao, M., Grossklags, J., & Chen, K. (2014). An exploratory study of white hat behaviors in a web vulnerability disclosure program. *Proceedings of the 2014 ACM Workshop on Security Information Workers*, 51–58. https://doi.org/10.1145/2663887.2663906

Zheng, Z., Xie, S., Dai, H.-N., Chen, X., & Wang, H. (2018). Blockchain challenges and opportunities: A survey. *International Journal of Web and Grid Services*, *14*(4), 352–375. https://doi.org/10.1504/IJWGS.2018.095647

Zimmerman, S. (2017). *Microsoft Announces Windows Bug Bounty Program and Extension of Hyper-V Bounty Program*. https://msrc.microsoft.com/blog/2017/07/announcing-the-windows-bounty-program/