Towards a Cyber Weapons Assessment Model – Assessment of the Technical Features of Malicious Software

Thomas Reinhold and Christian Reuter, *Science and Technology for Peace and Security (PEASEC), Technical University of Darmstadt,* Germany

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

***Index Terms*— Cyber weapons, cyberattack, classification, arms control, malware**

I. Introduction and Research Question

Over the last years, an increasing number of states have included cyberspace into their national security strategies [3] and their military planning [2]. A central element within these developments are "cyber weapons", the technical tools that can be used in the cyberspace for operations against foreign IT systems. Even the use of this term is controversial, because it has legal implications, especially in international humanitarian law, and so far, no internationally unified and binding definition exists. The concerns about an appropriate perspective on cyber weapons could easily be mistaken for a merely theoretical debate. Malware has been extensively researched and many important proposals for its classification and categorization have been made [4]. Nevertheless, an applicable and efficient definition of cyber weapons as a subset within the broad range of malware plays an essential role in the regulation of these destructive tools in international relations and the peaceful development of the cyberspace [6]. This is especially important for arms control measures such as export regulation, the prevention of unhindered proliferation [5] or treaties defining the dos and don'ts of the military application of such tools. Moreover, common criteria for these tools can further help to foster multilateral threat intelligence sharing platforms [7]. As this article shows, current approaches concerning definitions or classifications of cyber weapons are mostly either application- or actor-centric and concentrate on the intention or the deployment of malicious IT tools. These approaches perform sufficiently when applied after a specific incident but fail in situations where it is necessary to decide about the weapon character of a cyber tool prior to its usage. This is, at its core, the essential challenge of an effective restriction and monitoring of specific military cyber technologies [8].

A. Research Question and Methodology

This paper seeks to examine the following research question: **How can cyber weapons be differentiated within the complex and diverse landscape of malicious software based on features that are determinable without an assessment of their application context or any previous usage?** Our approach follows established arms control measures and looks for the critical components and thresholds that transform a technology or a specific item into a weapon. Such assessments have been established for non-cyber technologies over the last decades. This applies especially in the context of export controls, where manufacturers have to provide technical details on critical – potentially military – goods in order to get an export permission by assigned authorities. The authorities in turn analyze and compare these goods against thresholds and laws that have been defined and negotiated by international treaties and put into national legal norms. Our methodological approach is based on the finding, that after malicious cyber incidents, analytical assessments of the activities and especially the detected malware samples are carried out and published, that focus on the technical details

Thomas Reinhold is research associate and PhD student at PEASEC (Science and Technology for Peace and Security) at Technische Universität Darmstadt, Pankratiusstraße 2, 64289 Darmstadt, Germany (email: reinhold@peasec.tu-darmstadt.de)

Christian Reuter, Ph.D. is Full Professor and head of PEASEC (Science and Technology for Peace and Security) in the Department of Computer Science at Technische Universität Darmstadt, Universität Darmstadt, Pankratiusstraße 2, 64289 Darmstadt, Germany (email: reuter@peasec.tu-darmstadt.de)

such as code properties and capabilities, exploited vulnerabilities, applicated third-party libraries, similarities to existing malware samples, etc. Such technical details on cyber tools are – at least potentially – also available before their use and could therefore be assessed. Following this premise, our approach identifies the technical properties of the development and deployment of malicious software that can be measured "in situ". The findings are compiled into an assessment model for cyber weapons based on a set of analytical indicators as a foundation for arms control measures for the cyberspace. As a first step, Section II presents the range of existing approaches for the classifications of cyber weapons and highlights the research gap. Section III then analyzes and discusses the technical features of weaponized malware, identifying a set of measurable parameters and indicators. Building on this, Section IV provides our contribution, a suitable and practicable assessment model for cyber weapons, an explanation how this model can be applicated as well as an evaluation based on selected exemplary case studies. Section V concludes our approach by discussing the assessment model, its applicability and limitations for arms control measures, and presenting further research questions. The Annex will present selected technical details of the case study assessment.

II. Related Work: Current Approaches for the Classification of Cyber Weapons

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

A. Intent- and Effect-Based Classification

One of the initial approaches, which is still influential for current debates, has been provided by Rid and McBurney [9]. The authors dispense with any consideration of specific technical aspects of malicious code but instead focus only on the intention of the application as well as the deliberate selection of the targets by an attacker to distinguish malware from cyber weapons. Their concept already mentions that, besides their intention, malicious cyber tools can trigger unintended or even unforeseen consequences. This aspect of the triggered effects has been further elaborated by an approach of Brown and Tullos [10]. The authors propose a spectrum of the actual impact that ranges from non-invasive access and enabling operations, over non-destructive disruptions that suppress a service to destructive attacks. The authors consider the latter as cyberattacks and only the utilized software tools as cyber weapons. The most important approaches in the context of the categorization of cyber weapons were given by the two editions of the so-called Tallinn Manual [11], [12]. Cyber weapons are defined as tools within the cyberspace which are "*means of warfare that are by design, use, or intended use capable of causing either (i) injury to, or death of, persons; or (ii) damage to, or destruction of objects, that is, causing the consequences required for qualification of a cyber operation as an attack*". Although the Tallinn Manual was created by independent researchers for the NATO Cooperative Cyber Defense Centre of Excellence, it has become a quasi-official point of reference in international cyberspace politics and its perspective influenced many national security doctrines.

B. Classifications Based on the Strategic Assessments

Some researchers focused on the intent-based approach and the strategic perspective of the attacker and the circumstances of the attack. Mele [13] conceptualizes the concept of weapons by pointing out, that "*a weapon can be [...] an abstract concept thereby not necessarily a material one*". He proposes the consideration of both the strategic dimension – intended damage and the specific selection of a sensitive target – as well as the legal dimension – context and purpose – of a cyber operation. The author defines these as the "*typical elements of a cyber-weapon*". A similar premise is followed by Dewar [14], who criticizes that "*the subjectivity and context-dependence [...] causes particular difficulties when categorizing cyber tools as weapons*" because "*all weapons are tools, but not all tools are weapons*". To resolve this, the author urges to "*look at more conceptual issues regarding the incidents*" and to evaluate the assumed motivations of an attacker. As a conclusion he states that "*often the tool itself was not a digital device [...]. Techniques such as social engineering and phishing or the exploitation of unknown weaknesses in digital systems were the preferred tools*". Orye and Maennel [15] extend this assertion further by considering also the cognitive effects, which include "*sowing confusion, changing behavior, modifying trust, changing (public) opinion, manipulation*".

C. Classifications Based on Normative Aspects

An approach that analyzes the significance and impact of malicious cyber tools in international relations has been proposed by Stevens [16]. The author states that "*weapons can be understood as 'the violent materiality of the existential condition of uncertainty'*" and that cyber weapons "*shape a condition of marked uncertainty in the contemporary international order. Silent, invisible and potentially very effective, they are attractive to states and non-state actors seeking advantage in war and in peace*". This approach fundamentally questions the definition of cyber weapons in the context of global power and governance. An early attempt for regulation was undertaken with the extension of the Wassenaar Arrangement on Export Controls for Conventional Arms and Dual-Use Goods and Technologies [17]. In 2013, the multilateral treaty added the item of "*intrusion software*" to its list of regulated goods, with the following definition: "*Software specially designed or modified to avoid detection by 'monitoring tools', or to defeat 'protective countermeasures', of a computer or network-capable device, and performing any of the following: (a) The extraction of data or information, from a computer or network-capable device, or the modification of system or user data; or (b) The modification of the standard execution path of a program or process in order to allow the execution of externally provided instructions*." This approach defines cyber weapons by their potential capacity for malicious impact on IT systems according to the

effect-based classification.

D. Classifications Based on the Comparison with Traditional Weapons and Weapons Technology

Another attempt to define the nature of cyber weapons is based on the comparison with existing, well understood and examined weapon technologies. An early, but still important approach by Sommer and Brown [18] reflects the features of a generic kinetic weapon. In their point of view, a weapon is "*a directed force – its release can be controlled, there is a reasonable forecast of the effects it will have, and it will not damage the user, his friends or innocent third parties*". The authors highlight that "*there is an important distinction between something that causes unpleasant or even deadly effects and a weapon*". They suggest evaluating cyber tools based on these characteristics, in addition to the required usage capabilities, the target "*inside knowledge*", whether and how fast the tool can be detected before and during deployment, how quickly counter measures can be applied, and whether it is possible to detect and identify the perpetrator. An attempt by Hatch [19] states that the conditions for a cyber weapon are the system's fundamental design and initial consideration to "*act as a weapon*" and the "*capability to cause mass causalities at a single point in time and space*" like "*cyber operations that trigger a nuclear plant meltdown; open a dam above a populated area [...] or disable air traffic control services, resulting in airplane crashes*".

E. Classifications Based on Architectural Characteristics of Malicious Software

A few proposals have been made that focus on the architectural characteristics of malicious software. One major approach is presented by Herr [20] with the "*PrEP framework*", signifying "*propagation, exploits and payload*" as the base components of any malware that are required for a cyber weapon: "*The propagation method defines how the code is delivered into a target system and the payload is the core executable code of the malware that determines its functionality and delivers its effects. The [...] exploit, allows both propagation and payload delivery by taking advantage of vulnerabilities in computer systems and their defensive measures*". The author also explicitly criticizes the intent-based perspective, because "*intent and perception are diffuse characteristics and difficult to judge"*. The distinction between the different components of a malicious tool aims to provide an in-depth view on the specific technical elements that define a cyber weapon. A critical "*red line*" is specified in particular by "*a payload designed to create destructive physical or digital effects*", whereas the authors state that cyber weapons "*create physical and digital effects"* and that *"defining [cyber weapons] without them creates unhelpful limitations*". Contrary to the authors criticism of "*effect and intend*", this still includes the assessment of the actual impact and the anticipation of an attacker's aims into the cyberweapons assessment. Besides the important technical perspective, the approach does not further exemplify how to measure the capabilities of the payload, nor does it provide a structured and uniform analysis method of the malware.

Another technical approach by Maathuis et. al. [21] defines three different components or layers of malware: The access layer to reach and enter a foreign IT system and circumvent any defense mechanism, the transport layer for the propagation within given IT infrastructures, and the payload layer for the effect. The authors state that a cyber weapon is "*a computer program created and/or used to alter or damage [...] a system in order to achieve (military) objectives against adversaries inside and/or outside cyberspace*". They suggest the assessment of different additional technical, tactical, and strategic aspects of the malware like the configurability and adjustability, the sophistication, and the technical knowledge about the intended target as well as the disruptive or destructive intent and the selection of a relevant target.

F. Research Gap

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

III. TECHNICAL FEATURES OF CYBER WEAPONS

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

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

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

A. Production and Storage

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

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

In terms of the longevity of exploitable vulnerabilities, studies have shown the potentially enormous life span of up to nearly seven years until their detection and closure [22]. This problem is regularly confirmed by data breach reports [23], which show that most of the detected cyber incidents are based on already known vulnerabilities, which have not yet been patched in the targeted devices. Modular cyber weapons allow attackers to combine an existing payload with a vulnerability that matches the current target system. Therefore, the developed components of cyber weapons do not require any special maintenance in order to be reliable, apart from preventing other actors from gaining knowledge of these cyber weapon and its capabilities and establish safeguards for "*digital weapons arsenals*" in order to prevent any threats to the actor's own IT systems and the national IT security of states [24]. Table I summarizes these parameters.

TABLE I
PARAMETERS REGARDING THE PRODUCTION AND STORAGE

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

B. Availability and Steps for Full Operational Capability

The military deployment of weapon systems is usually targeted against specifically selected objects, which is often associated with an adjustment to the environment of the target and its vulnerabilities. In the case of cyber weapons, this preparation requires extensive knowledge of the object, information that can often only be gathered through reconnaissance operations, which potentially require hacking of minor, upstream IT systems. This highlights that an effective deployment relies on the capacities to circumvent all security measures on the way towards the target, including all upstream systems. Since these activities must remain hidden in order to prevent premature detection, the time required for deployment is also a decisive factor. Some military strategists argue that it is necessary or efficient to implant cyber weapons or to create hidden access possibilities in strategically relevant IT systems as a precautionary measure. The US Cyber Command extended this approach towards a "*persistent engagement*" [25], that includes the permanent deployment of cyber tools within adversary networks, forcing them to constantly observe, secure, and adapt their systems.

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

TABLE II
PARAMETERS REGARDING THE AVAILABILITY AND STEPS FOR FULL OPERATIONAL CAPABILITY

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

C. Deployment and Operation

As the deployment of cyber weapons takes place in multiple, consecutively triggered steps, such tools should be considered

using a shell model, like the basic three-layer approach proposed by Maathuis et al. mentioned above [21], or more complex models such as the cyber kill chain [26], a concept based on the work of Hutchins et al. [27]. Drawing from those, each shell should contain its own tools and capacities, whereas the actual configuration and required infrastructure depend on two parameters: First, the intended effect and impact, which can be tailored either to a specific target or to a class of IT systems or products. Second, the decision to what extent a deployed tool should be capable to propagate autonomously and trigger its payload. The possible options range from a "*fire and forget*" approach with an automatic operation based on built-in rules to a manually operated approach with command-and-control infrastructures that allow direct human control of the deployment process. Especially the chosen propagation mechanisms may limit the measures available to prevent unintended effects. Limiting an automatic infection to intended targets can be difficult to control, since the behavior of a particular code depends on the conditions of the actually infected system, which are often difficult to predict, possibly resulting in incalculable effects. Even without automatic payload activation, a widespread infection raises concerns about which of the infected systems are relevant to military goals. Even though these considerations are not dealt with in the context of this article, they are directly related to the "*human-in-the-loop*" debates that are highly controversial internationally in the field of lethal autonomous weapon systems [28] and the problems of meaningful human control [29]. A measure to prevent or stop the unintended propagation can be an explicitly built-in so-called "*kill switch*", a function that offers the possibility to completely shut down the operation of the cyber weapon. However, based on technical examination of detected malware samples, these functions are rarely used, as they are relatively easy to detect by the defenders, which undermines their effectiveness [30]. With regard to the penetration of IT systems, any unauthorized attempt to access an IT system can potentially damage that system. The circumvention of security and defense measures or the concealment of access from logging mechanisms manipulates the regular behavior of the system. Depending on the skills and expertise of the attacker and the information available about the attacked systems, this can have unintended or unexpected effects, leading to operational disruptions or system failures. In view of the international humanitarian norm of protection of civilian systems, this requires an assessment of each intruded IT system, what programs it runs and what external purpose it serves. These aspects are presented in the following table III.

TABLE III
PARAMETERS REGARDING THE DEPLOYMENT AND OPERATION

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

D. Direct Impact and Effects

The potential impact of cyber weapons can cover a broad spectrum, and the effective impact is strongly influenced by the weakness and vulnerability of the target which is reflected in the parameters of table IV. If a targeted system is not prone to a utilized vulnerability, has recently been updated and patched with security fixes, or has implemented strong IT security measures that detect and stop unusual system functions, a cyber weapon will either not be able to penetrate the target at all or will fail to launch its malicious payload. Other attack methods are based on the regular use of IT systems by overloading their processing capacity, which usually leads to their temporary shutdown. Although mitigation techniques exist to some extent [31], these attacks are very effective and are conceptually more difficult to prevent [32]. The different attack or infection approaches influence the possible reaction time of the defending actors regarding their chances of mitigating the effects and the malicious propagation, and thereby the effectiveness of the cyber weapon. A payload that has been secretly implanted into a target system limits the available defensive options, in contrast to cyberattacks that attempt to openly disrupt services. Once the payload has been triggered, the evolvement of its effects can vary widely depending on the configuration of the cyber weapon and the situation of the attacked system. It can range from a direct and contained impact on the targeted system (first level effects), impacts on connected IT systems that rely on certain services or functions of the targeted system (second level effects), to effects on other connected systems, either through propagation or chain effects (third level effects). The complete impact estimation contains a high potential for miscalculations or failures. The attacker needs to make assumptions about the target, its environment, dependencies, and the reaction of the attacked systems and actors while ensuring that the programming of the cyber weapon operates as expected and contains no errors. Nevertheless, a specific deployment may encounter unexpected conditions which can lead to a completely different effect or undesired effects.

The discussion on the participation of the China-based company Huawei in the construction of 5G mobile networks [33] highlights another aspect. Concerns have been expressed that malicious code could be directly integrated into widespread small off-the-shelf components, possibly granting unauthorized access or waiting for a trigger signal. In such cases it is extremely complicated to distinguish between the legitimate host system and the malicious code, especially when backdoors are suspected to be hard-wired into the chip design.

TABLE IV
PARAMETERS REGARDING THE IMPACT AND EVOLVEMENT OF EFFECTS

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

E. Overall Results, Successions and Leverage

A comprehensive assessment of the possible overall effects of a cyber weapon, which is presented in table V, is required

for the authorization of its application in light of the rules of international law such as the UN Charter [34]. For the current state of ongoing international debates, which have not yet been settled, the legal threshold is drawn at the point where the effects of a cyber weapon correlate with the "*use of force*" – usually interpreted as severe damage to objects or people – which is prohibited outside declared military conflicts [35]. A complete evaluation must also include the aftermath such as the reaction of the attacked and third-party actors. If the utilized malicious code uses zero-day exploits or other methods of intrusion unknown to the public, its usage reveals this secret, allowing defenders to adjust their protective measures. It also provides any other witnessing or later analyzing party with knowledge to learn and adapt, as long as the vulnerability is not completely fixed. This can result in threats to the attacker's own systems, as demonstrated by the EternalBlue vulnerability, which – originally owned by the NSA – was used in the malware campaign NotPetya [36] that also caused economic damage to industrial facilities in the US [37]. As already mentioned, neither the built-in logic of a tool nor a human conductor can safely and ultimately decide whether the penetrated system is a valid military target or not. The risk of mistakes is especially present during the intrusion, since at this point the attacked system can only be analyzed "*from the outside*". The potential effects on uninvolved systems and the risks of maloperation highlight that the actual effects caused by a cyber weapon can deviate considerably from its intention. A comprehensive assessment of such complex situations must therefore consider the following three dimensions to estimate the maximum possible effects:

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

TABLE V
PARAMETERS REGARDING RESULTS, SUCCESSIONS AND DAMAGES

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

IV. ASSESSMENT MODEL FOR CYBERWEAPONS AND CASE-STUDY-BASED EVALUATION

A. How to assess cyber weapons

The following section will propose an assessment model for cyber weapons that analyzes the capabilities of a given software and – given the intended application context of arms control – contrast the results with existing norms and regulations. The model cannot and does not aim to provide comparability between assessments – which would require any kind of scoring – but rather to present a structured method to assess specific features of a given software in order to provide a unified basis for the evaluation of its cyber weapon character. Therefore we propose a set of "cyber weapon indicators" as given in table IV, whereas each indicator is linked to multiple of the previously discussed technical parameters that relate to this indicator and connected with a possible range of expression. The indicator order follows the concept of the cyber kill chain [26], [27], an established method of malware life-cycle analysis that separates the different steps from the development of a malware to its deployment. Facilitating this order supports the unified assessment of the technical capabilities required for each step.

TABLE VI
INDICATORS FOR ASSESSMENT OF CYBER WEAPONS

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

The intended application of our proposed assessment model will stepwise assess each indicator by analyzing the associated parameters and the underlying questions regarding specific technical capabilities of the analyzed software. For the discussed context of arms control, this assessment will be performed by authorities like e.g. the German Federal Office of Economics and Export Control (BAFA - Bundesamt für Wirtschaft und Ausfuhrkontrolle) or in the U.S. Department of

Commerce's Bureau of Industry and Security (BIS) that are entitled and authorized to review and grant or deny export requests based on national laws and regulations. As already implemented for other technologies and goods, companies requesting an export license are required by law to provide technical documentation, source code samples, or compiled binaries of their products and submit these to the authorities. Taking into account that these documents and required information are probably not complete, the assessment results for each parameter can range between "yes", "no", "partially", and "unknown". This does provide neither a scoring nor binary answers, but taken together, its assessment allows to specify a position for each indicator within the provided range of expression. Although this leaves room for different considerations, the parameter assessments that focus on a specific capability and the question if a software contains it, provides in our opinion an appropriate degree of objectivity. Finally, the different indicator assessments in connection with the amount and distribution of "yes" and "partially" answers for the analyzed parameters provide the basis for a concluding decision on the cyber weapon character of the analyzed software. With regard to the intended application context, this decision will primarily be subject to legal regulations and political considerations. As usual for arms control and export regulation, the critical thresholds which technical capabilities are considered to manifest a weapon will likely differ for different states as long as no internationally binding norm or other treaties exist.

B. Case-study based evaluation

In order to evaluate the application of the identified indicators, the following section presents exemplary assessments, to find out whether the model applies to real-world cases. As there have been several incidents over the last years that have caused damage, we have chosen two "positive" cases that probably could be considered as cyber weapons and face these with one "negative" case, that is probably no cyber weapon in order to illustrate the contrast. With regard to the support decision character of our model, this is intended to be as an exemplification of its application, rather than a sufficient validation that would require systematic testing of cases that have been publicly classified as cyberweapons and cases that have not been labeled this way. This goes beyond the scope of this paper but is a task for future work. The intended use case of our model requires access to technical documentations and code samples, something that requires legal access possibilities for entitled authorities as these information are usually classified. To simulate this situation, we have chosen example cases of past incidents that have been provided with freely accessible analytical reports to test our assessment against the public assessment of the incident. The reports that we used have to focus on technical details of the malware such as reverse engineered and decompiled binaries, code samples, string analysis, comparison with known vulnerability and exploit databases, analysis of the propagation and communication capabilities. All technical information that we have taken into account should have been available to entitled arms control and export regulation authorities to this extent, even before the malware was used. We neither used knowledge of the malware outcome nor assumptions about the attacker's intentions. Based on a meta-analysis of the selected reports, the evaluation will assess the cases by testing the indicators stepwise and analyzing each associated parameter. To circumvent the current lack of any internationally binding legal definition, we facilitated the broadly approved Tallinn Manual perspective [11] for the final cyber weapon consideration in a slightly modified version where cyber weapons are tools that specifically contain the technical capability "[...] *of causing either (i) injury to, or death of, persons; or (ii) damage to, or destruction of objects*". The following subsections will briefly present the cases, the reports we used, our assessment, and finally the conclusion. In order to maintain the readability of the text, we dispense with single code examples in this section, but reference to selected examples of technical details and further descriptions that we present in the Annex of this paper. In addition, we list references and – if available – page numbers to the most relevant analytical findings and quotes of the reports to underline the technical foundation of our assessment. The detailed results for all indicators and assessed parameters are presented in table VII, where the assessment results are represented symbolically for a better overview with the following notation: "Yes" (●●), "No" (○○), "Partially" (●○) and "Unknown" (××).

1) First positive Case: Stuxnet

Although the Stuxnet incident dates back to the year 2010, it is presumably the best known and still the most thoroughly analyzed malware to date. It was discovered at the nuclear enrichment facility in Natanz, Islamic Republic of Iran and was used to achieve a beyond-the-normal wear of enrichment machines. Our assessment is primarily based on the analysis of Langers' "To kill a centrifuge" [1] and the highly technical Symantec reports on the initially detected [52] as well as an earlier version of Stuxnet [38]. The authors conclude that Stuxnet has been tailor-made as it e.g., had been manipulating the supervisory control room software that "*appears to be a genuine development for the Natanz Fuel Enrichment Plant*" [1, p.8] and contained exploits for its specific vulnerabilities [Annex A.1] to hide its activities as well as to intercept and manipulate the Step7 dubbed control of the industrial programmable logic controller (PLC) [Annex A.2] which regulates the actual industrial process. In addition, the code contained information on the specific configuration of the industrial hardware in this facility and "*infects [...] controllers with a matching configuration*" [1, p.8], [Annex A.3]. Stuxnet contained at least two different attack payloads, one that manipulated the rotor speed in centrifuges [1, p.10ff] and an earlier, yet much more dangerous version that can create overpressure in the enrichment devices [1, p.5ff], [38, p.9ff]: "*[it] contains an alternative attack strategy, closing valves within the uranium enrichment facility at Natanz, Iran, which would have caused serious damage to the centrifuges and uranium enrichment system as a whole*" [38, p.1]. Regarding the control of the attack, Stuxnet was able to develop a

communication channel despite the air-gapped system by facilitating infection and propagation methods that allowed to transfer information "*by compromising mobile computers of contractors who enjoy legitimate physical access to the target environment*" [1, p.22], [Annex A.4]. In contrast the authors conclude that "*there is no logic implemented in the malware which could actively disable the malicious code on infected controllers*" [1, p.18]. In line with these examples, the evaluation of all parameters draws a picture of a project that contains the capabilities to reach, attack, manipulate, and even destroy a specific IT system. The attacker exploited multiple zero-day vulnerabilities and had made significant efforts to avoid unintended side-effects or the detection of the attack and invested considerable know-how to establish a communication channel despite the limited direct controllability. Besides some several high-class exploits, Stuxnet contained no off-the-shelf utilities or code. Taking all of this into account, this assessment confirms the conclusion that Stuxnet has to be considered a cyber weapon.

2) Second positive Case: TRISIS/TRITON

The second example TRISIS/TRITON has been detected in 2017 in a petrochemical plant in Saudi Arabia. It presumably was manually injected and able to manipulate the Schneider Electric's Triconex safety instrumented system (SIS) that is responsible for reacting on critical operation incidents to deactivate the fail-safe operation of the industrial facility. For our analysis we used three reports from Dragos [39], FireEye [40], and the US National Cybersecurity and Communications Integration Center [41]. TRISIS code was designed to target a specific facility, identified by a SIS configuration that the malware was designed for: "*[As] each SIS is unique and to understand process implications would require specific knowledge of the process*" [39, p.3]. The malware required a manual injection to the facilities network [Annex B.1] and exploited a vulnerability of a specific version of the Triconex system [Annex B.2]. The malware contained code to perform different alternative attack methods [40, p.4] to manipulate or deactivate the SIS system "*that collectively would degrade industrial processes, or worse. Were both the process and the safety systems to be degraded simultaneously, persons, property, and/or the environment could suffer physical harm"* [41, p.18, p.12ff]. TRISIS code was written in Python and based on structure of the code, the possibilities to extend it with additional scripts it "*represents a facilitating capability or framework for the actual ladder logic change that has the potential [... to] be repurposed to deliver alternative payloads to either deliver different logic files (the external binaries uploaded by TRISIS to the target SIS) or to utilize differently embedded binaries to target different SIS types entirely."* [39, p. 13]. As TRISIS is built to be operated fully manually via hardcoded communication channels [38, p.12] the code contained "*anti-forensics technique to hide the presence of the attacker code on the Triconex controller"* [Annex B.3]. Regarding the complex and diverse infection, the capabilities for the evasion of security measures and the code injection process, the reports conclude that TRISIS development needed highly qualified, well-resourced adversaries with lengthy timelines [38, p.8]. This assessment of TRISIS shows that it contains capabilities to attack a strategically selected goal with methods for a rather long-term access and manipulation capabilities which could cause serious damage. This suggests the conclusion that TRISIS must also be considered a cyber weapon.

3) Negative Case: Emotet

The following subsection will present an incident, that - although it have caused serious damage - cannot considered to be cyber weapon according to our approach. This negative case example should illustrate the relevance of certain indicators and specific technical capabilities for our assessment and classification approach. The example we have chosen is Emotet, a trojan malware that was first detected in 2014 [49]. During its lifetime, the malware has been changed a lot and often been used as an preliminary infection step for multiple different malware campaigns that afterwards loaded additional payload code. According to the US-CERT, the malware is "*among the most costly and destructive malware affecting state, local, tribal, and territorial (SLTT) governments, and the private and public sectors*" [46]. At least one case is reported, where an infection directly impaired hardware by overheating [47]. With regard to this threat, the Emotet infrastructure has been took down in 2021 by an international coordinated action, led by EUROPOL [48]. For the technical assessment we used reports from Bromium [43], Malwarebytes [44], PaloAlto Networks [45] and the US Cybersecurity and Infrastructure Security Agency (CISA) [46]. According to these reports, Emotet is a flexible malware that facilitates a very "*effective combination of persistence and network propagation*" [44]. Over the years, the malware has emerged to an actual business, where "*the primary source of revenue for its operators may be through selling access to its botnet infrastructure to other malware operators, instead of directly monetizing stolen financial information*" [43, p.3]. Therefore, Emotet has developed to a broad toolbox that utilizes different phishing and watering hole infection methods: "*Emotet uses different techniques to distribute these [infected] Word documents. The malspam may contain an attached Microsoft Word document or have an attached ZIP archive containing the Word document. [...] Some emails distributing Emotet do not have any attachments. Instead, they contain a link to download the Word document. In previous years, malspam pushing Emotet has also used PDF attachments with embedded links to deliver these Emotet Word documents*" [45]. In addition, Emotet established a complex botnet infrastructure that is used, among other things, to deliver different payloads or download additional code from a set of an extendable set of modules for specific data grabbing and exfiltration tasks [50]. Although Emotet has been used for different campaigns, from simple ransomware attacks to attacks on more selected targets and state institutions like parts of the Lithuanian Government [51], its technical capabilities are focused on broadly executed phishing activities [Annex C.1], with a high degree of automated operation and a broad

impact. Even if targeted campaigns had facilitated custom target-relevant phishing emails or selected groups of email recipients, its technical capabilities are rather not built to be tailored made for a selected, certain ICT system [Annex C.2]. Besides the already mentioned different phishing methods of infected documents, this is also reflected in capabilities intended for spreading within networks: "*Once Trojan.Emotet has infected a networked machine, it will propagate by enumerating network resources and write to share drives, as well as brute force user accounts. Infected machines attempt to spread Emotet laterally via brute forcing of domain credentials, as well as externally via its built-in spam module*" [44]. In addition, Emotet did neither use zero-day exploits nor any unique target information like running service and its software versions and configuration, but rather exploited known vulnerabilities [50] of commonly used operating and office software with an "hit as much as you can" mentality and is even capable of injecting other malware into infected systems on a malware-as-a-service basis [Annex C.3]. These assessments lead to the conclusion, that on the one hand Emotet contains the capabilities for causing damage, but on the other hand lacks a specific target selection and tailoring as well as a manual steering of the attack, the payload delivery and its triggering. The code does not contain dedicated measures to prevent unintended proliferation and effects. Beside its indisputable destructive and dangerous nature, with regard to our assessment model, Emotet cannot be considered a cyber weapon and rather reminds of a tool for "digital vandalism" in the sense that it indiscriminately damages the things within its reach. This assessment is in line with the majority of Emotet's public perception.

C. Evaluation of results

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

TABLE VII
DETAILED EVALUATION OF SELECTED CASE STUDIES

Indicator	Case 1: Stuxnet		Case 2: TRISIS/TRITON		Case 3: Emotet	
	Evaluation	Assessment	Evaluation	Assessment	Evaluation	Assessment
Means of propagation	- Automatic, covered spreading over intruded networks - Propagation until a specific network (target) reached - Payload only activated within target - Injection, communication and control probably over an air gap or manipulated hardware - Specifically designed for Siemens SCADA product line "Step 7" and industrial hardware devices	P6: ●● P8: ●○ P10: ●●	- Potentially manual infection - Several different modules for privilege escalation and access - Specifically designed to infect Schneider Electric's "Triconex 3008 Safety Instrumented System" (SIS) controllers - Capability to manipulate failsafe behavior of industrial facility could have been used to force drastic damages with additional malware - Manual payload triggering and stopping	P6: ●● P8: ●● P10: ●●	- Multiple spreading mechanism, but following rather an "A lot helps a lot" approach for maximum proliferation - Capabilities to spread over networks quickly and establish backdoors for payload - No kill switch but similar IT security measures could be established temporarily by IT experts via exploiting a bug in Emotet	P6: ●○ P8: ○○ P10: ○○
Autonomy of deployment and application	- Intrusion, detection-prevention and propagation via built-in automatic routines for different Microsoft operation systems and Siemens software - High degree of concealment mechanisms - Monitoring, control, and software updates via different C2 servers and ad hoc P2P mechanism - Payload activation automatic based on built-in routines to detect the	P7: ●● P8: ●● P9: ●○ P10: ●● P14: ●○	- Manual deployment towards a specific system - Manual operation within target network - Custom made infection routines and customized process manipulating payload for a specific industrial facility and its SCADA architecture - Continuously adjusted concealment mechanisms	P7: ●● P8: ●● P9: ●● P10: ●● P14: ●○	- Broad, automated deployment - No manual operation - Capabilities for hit-and-run campaigns with broad infection and payload propagation - Later additional payload download possible but no manual steering on single infected systems - Common obfuscation and encryption of malware files to prevent automated AV countermeasures - No kill switch	P7: ●○ P8: ○○ P9: ●○ P10: ○○ P14: ●○

	intended target, anticipating the air gap of the target				- Botnet infrastructure for data exfiltration and additional payload provision	
Controllability and intervention measures	- Intended, but faulty automatic disablement of propagation via time settings of infected system - No dedicated "kill switch" - Probably a dedicated testing facility of industrial target - Limited communication with C2 servers, mostly built-in, but updatable	P3: ●● P5: ×× P9: ●○ P13: ●●	- Direct access to infected system and human controlled operation - Payload tailored based on situational conditions - Probably a dedicated testing facility of industrial target - Payload tested on infected device before finally deployed	P3: ●● P5: ×× P9: ●● P13: ●●	- Relatively "open" infra-structure meant to be operated by third parties - Custom payload possible, but not for single infected systems - "Targeted" only in the sense of dedicated email recipients and tailored phishing emails - No proliferation containment	P3: ×× P5: ○○ P9: ●○ P13: ○○
Required infrastructures	- Malware itself independent from C2 infrastructures - Communication and data exchange channels via different, separated measures	P1: ●● P7: ●○ P8: ●● P14: ●○	- C2 infrastructures for deployment and operation, Hardcoded DNS servers - Communication and data exchange channels via different measures - Extendable via loadable code modules	P1: ●● P7: ●● P8: ●● P14: ●○	- Modular software with different infection and persistence methods - Custom payload with option for later download - C2 botnet infrastructure for payload provision and data extraction	P1: ●● P7: ●○ P8: ○○ P14: ●○
Quality of penetration measures	- Different modules and measures for penetration and propagation - Multiple 0day exploits - Bridging the air gap - Redundant measures for application life cycle (infection, data drop off, communication) supporting the autonomous propagation and infection over different systems - Built-in extensive knowledge of target environment and vulnerability	P1: ●● P2: ●● P4: ●● P12: ●○	- Extendable, continuously refactored Framework architecture - Payload independent from interchangeable initial infection and persistence capabilities - Payload injection and operational code tailored for specific devices - Built-in extensive knowledge of target environment and vulnerability - If combined with harmful payload, immediate physical effects possible	P1: ●● P2: ●● P4: ●● P12: ○○	- Continuously extended software basis - Exchangeable infection, propagation and persistence methods - Independent payload - Not developed to reach and infect single target systems and exploit their specific vulnerabilities - Developed for broad, quick and effective infections and backdoor establishment - No zero-day exploits - Optional immediate payload triggering	P1: ●● P2: ○○ P4: ●○ P12: ●○
Direct payload effect	- Payload explicitly developed for a specific software version and production line of industrial hardware - Interchangeable Payload - Propagation mechanism developed to reach a specific target via multiple, different measures - Different measures for direct impact from direct immediate harm (v0.5) to slow sabotage (v1.0) - No direct defending possibility, but system shutdown	P2: ●● P4: ●● P6: ●● P12: ●○	- Payload explicitly developed for a specific software version and production line of industrial hardware - Payload interchangeable - By manipulating the failsafe behavior of the facility, direct harmful impact with additional malware possible without defending or mitigating possibilities - Manual operation allows to prevent collateral infections and payload deployment	P2: ●● P4: ●● P6: ●● P12: ○○	- Optional immediate payload triggering - Payload interchangeable and option for later download after backdoor established - Third party payload injection possible - No manual operation - No zero-day exploits used - Multiple spreading mechanism, following rather a "A lot helps a lot" approach for maximum proliferation - No actively harming payloads known	P2: ○○ P4: ●○ P6: ●○ P12: ●○
Unintended effects	- Presumably a high level of diligence by testing in a dedicated testing facility and replacement of Stuxnet v0.5 payload - Automatic propagation, but malicious payload triggered only on the target system - Integrated, though faulty "kill switch" - Precise impact - Potential spread of zero-day exploits	P3: ●● P5: ×× P9: ●● P10: ●● P11: ●● P13: ●○ P15: ●● P16: ●●	- Presumably a high level of diligence by testing in dedicated testing facility - Manual operation based on direct feedback prevented deployment and payload errors - Proliferation of new, highly critical attack vectors for SCADA systems - Uncalculatable destructive effects if combined with destructive malware	P3: ●● P5: ×× P9: ●● P10: ●● P11: ●● P13: ●● P15: ●● P16: ●●	- Automated, uncontainable propagation and infection - "Targeted" only in the sense of dedicated email recipients and tailored phishing emails - Optional automated payload triggering - No manual steering of payload triggering on single infected systems - Third party payload possible - No kill switch - No zero-day exploits	P3: ×× P5: ○○ P9: ●○ P10: ○○ P11: ○○ P13: ○○ P15: ○○ P16: ○○
Legend: The assessment results are symbolically represented with the following notation: "Yes" (●●), "No" (○○), "Partially" (●○) and "Unknown" (××)						

V. Conclusion and Future Work

A. The Technical Assessment of Cyber Weapons

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

B. Limitations

The assessment model with the proposed list of indicators does not claim to be exhaustive. It is rather intended to provide a standardized and unified procedure to determine if a specific malware can be considered a cyber weapon. In addition, the indicators can be utilized to cluster specific weaponizable functionalities of malware that characterize such weaponizable tools. Such a generalization, that considers a broad range of parameters, cannot provide "*sharp edges*" as dual-use aspects or incomplete information will influence decisions. The proposed approach is therefore optimized as a decision support for case-by-case assessments.This reflects also the limited area of application, as detailed technical documents, code samples or other technical information on the software are necessary. Therefore, the legal and institutional foundation to request and assess this information, e.g. as part of export control regimes or in the context of a vulnerability equity process [54], as well as their sensitive and probably secret nature needs to be considered when conducting our proposed assessment. The completeness of the available information also directly influences the amount and the certainty of assessable parameters. Nevertheless, these specific requirements and the political will are given for the intended context of arms control and its application via entitled authorities that are legally allowed to request the required technical details.

C. Further Research

The indicators and parameters identified can provide applicable measures for evaluating the cyber weapon character of malicious cyber tools. In addition to case-by-case decisions, they can also be used to cluster existing malware based on their technological approaches and capabilities. Further research should explore the development of a deterministic indication algorithm that combines the indications to weighted numerical values in order to compare different tools as well as to establish decision thresholds. In addition, a systematic study of more past incidents could support this refinement alongside a validation of the indicators as well as a possible identification of edge cases that need to be considered. Besides the task of the identification of cyber weapons, the analysis shows that the risks of unintended effects are high and depend on many aspects of the target system, some of which are difficult to assess. Further research can refine the definition of minimum considerations and implementation principles that help to minimize the risk of unintended effects, in line with international humanitarian law and its prohibition to attack "*objects indispensable to the survival of the population*" [42]. Finally, the ongoing militarization of cyberspace, with its consequences for international security, require a substantial, non-commercially motivated involvement of the computer sciences and a commitment to political issues, as many political challenges of cyberwar and its prevention have a deep rooting in technical details. This affects the development of technical measures for cyber arms control and its non-proliferation, the assessment of cyber attack methods, or the question how military cyber activities could follow international human rights rules, such as the distinction between civilian and military objects in the cyberspace. An understanding of these technical challenges, the stronger cooperation between computer sciences and politics, and the "translation" between these domains may pave the necessary way towards a stable and secure global cyberspace.

Annex

This Annex presents selected examples of technical details regarding the assessment of the three selected cases from section IV.B.

A. Stuxnet

1. List of exploited zero-day vulnerabilities for all detected Stuxnet versions [38, p.2]
   - MS09-025
   - CVE-2010-2568
   - CVE-2010-2772
   - CVE-2012-3015
   - CVE-2008-4250
   - CVE-2010-2729
   - CVE-2010-2743
   - CVE-2010-3888
2. PLC-Step7 Communication Manipulation
   "*The Step7 software uses a library file called s7otbxdx.dll to perform the actual communication with the PLC. The Step7 program calls different routines in this .dll file when it wants to access the PLC. For example, if a block of code is to be read from the PLC using Step7, the routine s7blk_read is called. The code in s7otbxdx.dll accesses the PLC, reads the code, and passes it back to the Step7 program. Stuxnet [...] renames the original s7otbxdx.dll file to s7otbxsx.dll. It then replaces the original .dll file with its own version. Stuxnet can now intercept any call that is made to access the PLC from*

any software package" [52, p.37]

3. Target checking and selection process
   Stuxnet searches for an industrial plant from Siemens with a specific hardware configuration by searching in the code for "*symbol labels [that] loosely follow the ANSI/ISA S5.1 Instrumentation Symbols and Identification standard used in Piping and Instrumentation Diagrams*" [38, p.6]. Each PLC is identified by a fingerprint label following this format: "`<delimiter><FunctionIdentifier><delimiter> <CascadeModule><delimiter><CascadeNumber> <DeviceNumber>`" [38, p.6]. Since the concrete PLC configuration of an industrial plant is unique, Stuxnet checked the amount and type of all PLCs it detected and compared this against a built-in list of PLC fingerprints to identify a specific industrial facility.
4. Jumping the Air gap via removable drive infections
   "*Stuxnet will copy itself and its supporting files to available removable drives any time a removable drive is inserted, and has the ability to do so if specifically instructed" [52, p.29ff], thus e*xploiting the LNK vulnerability CVE-2010-2568. "*Stuxnet will first verify it is running within services.exe, and determines which version of Windows it is running on. Next, it creates a new hidden window with the class name 'AFX64c313' that waits for a removable drive to be inserted (via the WM_DEVICECHANGE message), verifies it contains a logical volume (has a type of DBT_DEVTYP_VOLUME), and is a removable drive (has a drive type of DEVICE_REMOVABLE).*" After checking if the drive is suitable, "*.lnk files are created using Resource 240 as a template and four are needed as each specifically targets one or more different versions of Windows including Windows 2000, Windows XP, Windows Server 2003, Windows Vista, and Windows 7. The .lnk files contain an exploit that will automatically execute ~WTR4141.tmp when simply viewing the folder.*" [52, p.29ff] to inject Stuxnet into the system processing, allowing its hidden operations.

B. TRISIS/TRITON

1. *Target discrimination and spear headed design*
   "TRISIS is a Stage 2 ICS Attack capability, as defined by the ICS Cyber Kill Chain (...). Given its design and assessed use, TRISIS has no role or applicability to IT environments and is a focused ICS effects tool. As a result, TRISIS' use and deployment requires that an adversary has already achieved success in Stage 1 of the ICS Cyber Kill Chain" – Identifying and gaining access to a system able to communicate with target SIS – "and either compromised the business IT network or has identified an alternative means of accessing the ICS network. Once in position, the adversary can deploy TRISIS on its target: an SIS device." [39, p.9]
2. Exploited vulnerabilities for privilege escalation
   Triton leverages a "*previously-unknown vulnerability affecting Tricon MP3008 firmware versions 10.0–10.4 [that] allows an insecurely-written system call to be exploited to achieve an arbitrary 2-byte write primitive, which is then used to gain supervisor privileges*." Regarding the output addresses of the exploited system call "*No checking is performed (...) to ensure the pointers do not refer to the firmware region or other protected areas. This allows for data to be written to normally immutable and privileged regions.*" [41, p.15-16]
3. Anti-forensic and evasion techniques
   As Triconex and the SIS systems are highly safety-critical, they contain numerous fail-safe techniques, like checksum comparisons to ensure the validity of the code. Triton contained a dedicated module *crc.py within its* loaded *library.zip* of compiled Python modules that "*implements or imports a number of standard Cyclic Redundancy Check (CRC) functions*" [39, p.7] that are used to patch a "s*pecific RAM/ROM consistency check*" in order to "*prevent a fault from occurring when the firmware region does not match the ROM image that was loaded. Without patching this check, the injector would not be able to write the payload into the firmware region or modify the jump table to point to it without faulting the device.*" [39, p.14]. Additionally, "*after payload files were inserted into memory on the Triconex controller, the script initiated a countdown, periodically checking the status of the controller. If an error was detected, the communication library's method SafeAppendProgramMod attempted to reset the controller to the previous state using a TriStation protocol command. If this failed, trilog.exe attempted to write a small 'dummy' program to memory. We assess that this was an anti-forensics technique to hide the presence of the attacker code on the Triconex controller" [40].*

C. Emotet

1. Phishing and data breaching variations
   "*Emotet uses five known spreader modules: NetPass.exe, WebBrowserPassView, Mail PassView, Outlook scraper, and a credential enumerator" [46]. These* different tools can be used independently by loading different payload files into the memory once the victim is infected and information about the system are sent back to the C2 servers. The payload ranges from functions to collect passwords and user credentials from infected systems and external drives, read email addresses from Outlook accounts, send phishing mails, collect passwords and credentials from different web browser storage files, fetch "*passwords and account details for various email clients such as Microsoft Outlook, Windows Mail, Mozilla Thunderbird, Hotmail, Yahoo! Mail, and Gmail" [46] and* query network resources for further vulnerable systems.
2. Variations of initial infection mechanisms
   *Emotet*s first infection step is the spreading via different spam campaigns that lure the victim into downloading the malware. *"The email content may have a malicious link leading the victim to the Emotet downloader, or in other cases the downloader is delivered as the email attachment. We have seen MS Office Word documents, Excel spreadsheets, PDFs, JavaScript, and even password-protected ZIP files as the attachment. The most highly evolved spamming method, which appeared in*

recent months, is when the malicious object is inserted into a legitimate email conversation thread" [53]. In each case, malicious code is loaded from a range of different C&C servers either via direct download, VBA macros, or MS Windows shell functions.

3. Malware-as-a-Service capabilities

Beside the malware's own payload files, "*Emotet has the ability to install other malware and to infect the machine with it. There are examples where it has distributed other banking trojans including Qbot, Dridex, Ursnif/Gozi, Gootkit, IcedID, AZORult and Trickbot and then ransomware such as Ryuk, BitPaymer or MegaCortex. In cases where additional malware is delivered besides the modules, the executeFlag in the response is set to 0x03, leading the delivered malware to the 'C:\ProgramData' folder with a randomly generated name. I have seen a downloaded Ursnif variant with a list of the most common latest modules. It injected control.exe under the 'C:\Windows\System32' directory, which further injected code into explorer.exe. It copied itself to the '%APPDATA%\Microsoft\[random]' folder and set the AutoRun registry to gain persistence*" [53].

REFERENCES

[1] R. Langner, "To Kill a Centrifuge - A Technical Analysis of What Stuxnet's Creators Tried to Achieve ," 2013, [Online]. Available: http://www.langner.com/en/2013/11/20/langner's-final-stuxnet-analysis-comes-with-surprises/.

[2] S. Zeadally and A. Flowers, "Cyberwar: The what, when, why, and how [Commentary]," *IEEE Technology and Society Magazine*, vol. 33, no. 3. Institute of Electrical and Electronics Engineers Inc., pp. 14–21, Sep. 01, 2014, doi: 10.1109/MTS.2014.2345196.

[3] UNIDIR, "The Cyber Index - International Security Trends and Realities." 2013, [Online]. Available: http://www.unidir.org/files/publications/pdfs/cyber-index-2013-en-463.pdf.

[4] Y. Ding, R. Wu, and X. Zhang, "Ontology-based knowledge representation for malware individuals and families," *Comput. Secur.*, vol. 87, p. 101574, Nov. 2019, doi: 10.1016/J.COSE.2019.101574.

[5] K. Brockmann, "Challenges To Multilateral export Controls. The Case for Inter-regime Dialogue and Coordination." SIPRI, 2019, [Online]. Available: https://www.sipri.org/publications/2019/other-publications/challenges-multilateral-export-controls-case-inter-regime-dialogue-and-coordination.

[6] L. J. Robertson, A. Munoz, and K. Michael, "Managing Technological Vulnerability of Urban Dwellers: Analysis, Trends and Solutions," *IEEE Trans. Technol. Soc.*, vol. 1, no. 1, pp. 48–59, 2020, doi: 10.1109/TTS.2020.2975806.

[7] T. D. Wagner, K. Mahbub, E. Palomar, and A. E. Abdallah, "Cyber threat intelligence sharing: Survey and research directions," *Comput. Secur.*, vol. 87, p. 101589, Nov. 2019, doi: 10.1016/J.COSE.2019.101589.

[8] T. Reinhold and C. Reuter, "Arms Control and its Applicability to Cyberspace," in *Information Technology for Peace and Security - IT-Applications and Infrastructures in Conflicts, Crises, War, and Peace*, Wiesbaden: Springer Fachmedien Wiesbaden, 2019, pp. 207–231.

[9] T. Rid and P. McBurney, "Cyber-Weapons," *RUSI J.*, vol. 157, no. 1, pp. 6–13, Feb. 2012, doi: 10.1080/03071847.2012.664354.

[10] G. D. Brown and O. W. Tullos, "On the Spectrum of Cyberspace Operations," *Small Wars J.*, 2012, [Online]. Available: http://smallwarsjournal.com/jrnl/art/on-the-spectrum-of-cyberspace-operations.

[11] M. Schmitt, "The Tallinn Manual on the International Law Applicable to Cyber Warfare." Cambridge University Press, Cambridge, 2013, doi: 10.1017/CBO9781139169288.

[12] M. N. Schmitt and L. Vihul, "Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations." Cambridge University Press, Cambridge, 2017, doi: 10.1017/9781316822524.

[13] S. Mele, "Cyber-weapons: legal and strategic aspects." Italian Institute of Strategic Studies "Niccolò Machiavelli", Roma, 2013.

[14] R. Dewar, "Cyberweapons: Capability, Intent and Context in Cyberdefense," *CSS Cyber Def. Trend Anal.*, vol. 2, 2017, doi: 10.13140/RG.2.2.25947.05923.

[15] E. Orye and O. M. Maennel, "Recommendations for Enhancing the Results of Cyber Effects," pp. 1–19, 2019, doi: 10.23919/cycon.2019.8756649.

[16] T. Stevens, "Cyberweapons: power and the governance of the invisible," *Int. Polit.*, vol. 55, pp. 482–502, Jun. 2017, doi: 10.1057/s41311-017-0088-y.

[17] Wassenaar, "Recommendations for the Implementation of the 2013 Wassenaar Arrangement Changes Regarding 'Intrusion Software' and 'IP Network Communications Surveillance Systems,'" 2014, [Online]. Available: http://oti.newamerica.net/sites/newamerica.net/files/articles/Joint_Recommendations_Wassenaar_Implementation.pdf.

[18] P. Sommer and I. Brown, "OECD Study - Reducing Systemic Cybersecurity Risk," *OECD/IFP Proj. "Future Glob. Shock.*, p. 121, 2010.

[19] B. B. Hatch, "Defining a Class of Cyber Weapons as WMD: An Examination of the Merits," *J. Strateg. Secur.*, vol. 11, pp. 43–61, Jun. 2018, doi: 10.5038/1944-0472.11.1.1657.

[20] T. Herr, "PrEP: A Framework for Malware and Cyber Weapons," *J. Inf. Warf.*, vol. 13, pp. 87–106, 2014, doi: 10.2307/26487013.

[21] C. Maathuis, W. Pieters, and J. Van Den Berg, "Cyber weapons: a profiling framework," in *2016 International Conference on Cyber Conflict (CyCon U.S.)*, Oct. 2016, pp. 1–8, doi: 10.1109/CYCONUS.2016.7836621.

[22] L. Ablon and A. Bogart, "Zero Days, Thousands of Nights: The Life and Times of Zero-Day Vulnerabilities and Their Exploits", RAND Corporation, Santa Monica, 2017.

[23] Verizon, "2019 Data Breach Investigations Report," *Verizon Bus. J.*, 2019, [Online]. Available: file:///C:/Users/Edward S. Forde/Downloads/rp_Verizon-DBIR-2014_en_xg.pdf%5Cnhttp://www.verizonenterprise.com/resources/reports/rp_Verizon-DBIR-2014_en_xg.pdf.

[24] T. Reinhold and M. Schulze, "Wannacry About the Tragedy of the Commons? Game-Theory and the Failure of Global Vulnerability Disclosure," *Proc. 17th Eur. Conf. Cyber Warf. Secur. ECCWS 2018, pp. 454–463*, 2018, [Online]. Available: http://www.academic-bookshop.com/ourshop/prod_6457309-ECCWS-2018-PDF-Proceedings-of-the-17th-European-Conference-on-Cyber-Warfare-and-Security.html.

[25] USA-DOD, "Achieve and Maintain Cyberspace Superiority Command Vision for US Cyber Command," pp. 1–12, 2018, [Online]. Available: https://www.cybercom.mil/Portals/56/Documents/USCYBERCOM Vision April 2018.pdf?ver=2018-06-14-152556-010.

[26] B. Wrozek, "Cyber Kill Chain Methodology," 2017. Accessed: Jun. 07, 2019. [Online]. Available: http://m.isaca.org/chapters3/Charlotte/Events/Documents/Event Presentations/12062017/Cyber_Kill_Chain_Wrozek.pdf.

[27] E. Hutchins, M. Cloppert, and R. Amin, "Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains," *6th Int. Conf. Inf. Warf. Secur. ICIW 2011*, no. July 2005, pp. 113–125, 2011.

[28] UN-GGE, "Draft Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems." Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, 2019, [Online]. Available: https://www.unog.ch/80256EDD006B8954/(httpAssets)/6B80F9385F6B505FC12581D4006633F8/$file/2017_GGEonLAWS_WP9_Switzerland.pdf.

[29] M. Cannellos and R. Haga, "Lost in Translation: Getting Autonomous Weapons Systems Ethicists, Regulators, and Technologists to Speak the Same Language," *IEEE Technol. Soc. Mag.*, vol. 35, no. 3, pp. 50–58, Sep. 2016, doi: 10.1109/MTS.2016.2593218.

[30] Symantec, "Internet Security Threat Report. Ransomware 2017," 2017. Accessed: Jun. 16, 2019. [Online]. Available: https://www.symantec.com/content/dam/symantec/docs/security-center/white-papers/istr-ransomware-2017-en.pdf.

[31] O. Osanaiye, K. K. R. Choo, and M. Dlodlo, "Distributed denial of service (DDoS) resilience in cloud: Review and conceptual cloud DDoS mitigation framework," *Journal of Network and Computer Applications*,

vol. 67. Academic Press, pp. 147–165, May 01, 2016, doi: 10.1016/j.jnca.2016.01.001.
[32] A. Lavrenovs, "Towards Measuring Global DDoS Attack Capacity," in *2019 11th International Conference on Cyber Conflict (CyCon)*, 2019, pp. 1–15, doi: 10.23919/cycon.2019.8756851.
[33] D. Rupprecht, A. Dabrowski, T. Holz, E. Weippl, and C. Popper, "On security research towards future mobile network generations," *IEEE Commun. Surv. Tutorials*, vol. 20, no. 3, pp. 2518–2542, 2018, doi: 10.1109/COMST.2018.2820728.
[34] B. Sander, "The Sound of Silence: International Law and the Governance of Peacetime Cyber Operations," pp. 1–21, 2019, doi: 10.23919/cycon.2019.8756882.
[35] J. Kosseff, "The Contours of 'Defend Forward' Under International Law," in *11th International Conference on Cyber Conflict (CyCon)*, 2019, pp. 1–13, doi: 10.23919/cycon.2019.8757141.
[36] ESET, "One year later: EternalBlue exploit more popular now than during WannaCryptor outbreak," *ESET*, 2018. https://www.welivesecurity.com/2018/05/10/one-year-later-eternalblue-exploit-wannacryptor/ (accessed Jun. 16, 2019).
[37] Maersk, "Maersk improves underlying profit and grows revenue in first half of the year." Maersk, 2017, [Online]. Available: https://www.maersk.com/news/2018/06/29/20170816-a-p-moller-maersk-improves-underlying-profit-and-grows-revenue-in-first-half-of-the-year.
[38] Symantec, "Stuxnet 0.5: The Missing Link", Symantec 2013, [Online]. Available: https://docs.broadcom.com/doc/stuxnet-missing-link-13-en
[39] Dragos Inc., "TRISIS Malware," pp. 1–19, 2017, [Online]. Available: https://www.energy.senate.gov/public/index.cfm/files/serve?File_id=40B2ED59-D34E-47C3-B9E2-1E8D030C5748.
[40] B. Johnson, D. Caban, M. Krotofil, D. Scali, N. Brubaker, and C. Glyer, "Attackers Deploy New ICS Attack Framework 'TRITON' and Cause Operational Disruption to Critical Infrastructure," *Fireeye Threat Res.*, 2017, [Online]. Available: https://www.fireeye.com/blog/threat-research/2017/12/attackers-deploy-new-ics-attack-framework-triton.html.
[41] NCCIC, "Malware Analysis MAR-17-352-01 HatMan — Safety System Targeted Malware (Update B)," *Cybersecurity Infrastruct. Secur. Agency*, pp. 1–23, 2019, [Online]. Available: https://www.US-CERT.gov/sites/default/files/documents/MAR-17-352-01 HatMan - Safety System Targeted Malware %28Update B%29.pdf.
[42] IRCR, "Cyber operations and international humanitarian law: five key points." ICRC, 2019, [Online]. Available: https://blogs.icrc.org/law-and-policy/2019/11/28/cyber-operations-ihl-five-key-points/.
[43] Bromium, "Emotet: a technical analysis of the destructive, polymorphic malware.", Bromium Inc., 2019, [Online]. Available: https://www.bromium.com/wp-content/uploads/2019/07/Bromium-Emotet-Technical-Analysis-Report.pdf
[44] Malwarebytes, "Trojan.Emotet.", Malwarebytes Labs, 2020, [Online]. Available: https://blog.malwarebytes.com/detections/trojan-emotet/
[45] B. Duncan, "Wireshark Tutorial: Examining Emotet Infection Traffic", PaloAlto Networks, 2021, [Online]. Available: https://unit42.paloaltonetworks.com/wireshark-tutorial-emotet-infection/
[46] US-CERT, "Alert (TA18-201A) - Emotet Malware", CISA, 2018, [Online]. Available: https://US-CERT.cisa.gov/ncas/alerts/TA18-201A
[47] S. Gatlan, "Microsoft: Emotet Took Down a Network by Overheating All Computers", Bleeping Computer, 2020, [Online]. Available: https://www.bleepingcomputer.com/news/security/microsoft-emotet-took-down-a-network-by-overheating-all-computers/
[48] EUROPOL, "World's Most Dangerous Malware Emotet Disrupted Through Global Action", EUROPOL, 2021, [Online]. Available: https://www.europol.europa.eu/newsroom/news/world%E2%80%99s-most-dangerous-malware-emotet-disrupted-through-global-action
[49] F. Axel et. al., "A Comprehensive Look at Emotet's Summer 2020 Return", PROOFPOINT, 2020, [Online]. Available: https://www.proofpoint.com/us/blog/threat-insight/comprehensive-look-emotets-summer-2020-return
[50] US-CERT, "Alert (AA20-280A) - Emotet Malware", CISA, 2020, [Online]. Available: https://US-CERT.cisa.gov/ncas/alerts/aa20-280a
[51] BNS/TBT Staff, "Several institutions affected by email virus in Lithuania – center", The Baltic Times, 2020, [Online]. Available: https://www.baltictimes.com/several_institutions_affected_by_email_virus_in_lithuania___center/
[52] N. Falliere, L. O. Murchu, and E. Chien, "W32.Stuxnet Dossier V1.4", Symantec, 2011, [Online]. Available: https://www.wired.com/images_blogs/threatlevel/2010/11/w32_stuxnet_dossier.pdf.
[53] L. Nagy, "VB2019 paper : Exploring Emotet , an elaborate everyday enigma A brief history of Emotet", Virus Bulletin, 2019, [Online]. Available: https://www.virusbulletin.com/virusbulletin/2019/10/vb2019-paper-exploring-emotet-elaborate-everyday-enigma/.
[54] M. Schulze, "Quo Vadis Cyber Arms Control? – A Sketch of an International Vulnerability Equities Process and a 0-Day Emissions Trading Regime", Proceedings of the Interdisciplinary Conference on Technical Peace and Security Research 2019, pp. 24-39, 2019

Thomas Reinhold is research associate and PhD student at PEASEC (Science and Technology for Peace and Security) at Technische Universität Darmstadt. He has been a non-resident fellow at the Institute for Peace Research and Security Policy at the University of Hamburg since 2009. Concerned with social effects of technology, his field of work are the threats in cyberspace and its increasing militarization and issues of disarmament and arms control. Since 2017 he has been a member of the Transatlantic Cyber Forum and the Research Advisory Group of the Global Commission on the Stability of Cyberspace (GCSC).

Christian Reuter, Ph.D. is Full Professor at Technical University of Darmstadt. He holds a Diplom, M.Sc. and Ph.D. in Information Systems. His chair Science and Technology for Peace and Security (PEASEC) combines computer science with peace and security research. On the intersection of the disciplines cyber security and privacy, peace and conflict studies as well as human-computer interaction, he and his team specifically address (1) peace informatics and technical peace research, (2) crisis informatics and information warfare as well as (3) usable safety, security and privacy.