Verification in Cyberspace

11

Thomas Reinhold and Christian Reuter

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 treaty compliance by observing, counting or monitoring specific actions and their accordance with the agreed 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 chemical weapons technology. Yet, cyberspace is a human-made domain; adjusting its technical setting, rules, and principles may help 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.

T. Reinhold (✉) · C. Reuter
Science and Technology for Peace and Security (PEASEC),
Technische Universität Darmstadt, Darmstadt, Germany
e-mail: reinhold@peasec.de

C. Reuter
e-mail: reuter@peasec.tu-darmstadt.de

C. Reuter (ed.), *Information Technology for Peace and Security*,
Technology, Peace and Security I Technologie, Frieden und Sicherheit,
https://doi.org/10.1007/978-3-658-44810-3_11

Objectives

- Understanding the concept of verification in the context of international security politics as well as examples of verification for currently existing military technologies.
- Identifying the technical features of cyberspace that hinder the development of verification measures for this domain.
- Gaining insight into how verification measures for cyberspace need to work, which technical features of this space can be used for measures and checks, and which established IT approaches and methods from other areas could be applied to develop such measures.

11.1 What is Verification?

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 "*pacta sunt servanda*" (Wehberg, 1959), which 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). This raises the question of which instance should be in charge of checking the compliance of states with specific treaties and how this should be performed. This question has 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. This is mainly ruling out the possibility of higher instances. Therefore, states often regulate their relations by mutual agreements. A complementary tool for treaties is the possibility of treaty partners checking each other's compliance 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[1] that thereby can demonstrate their good intentions. Verification is an important measure for international security politics and is mainly integrated into so-called **verification regimes**, a concept that is based on the regime theory of Robert O. Keohane (Keohane & Martin, 1995). 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

[1] 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).

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 belong to rules the treaty members agree to follow.
- The practical measures that treaty members or specifically entrusted authorities are allowed to perform to check the treaty members' compliance.
- Optionally, the definition of the authority 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 have been developed over the last decades for different reasons and situations. They are based on different mandates, often in the context of disarmament, arms control or so-called **non-proliferation**[2] 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 famous example of 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)[3], the IAEA has been put into charge of 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).

[2] 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.

[3] 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).

One of its most popular tasks was to check Iran's compliance with the JCPOA (*Joint Comprehensive Plan of Action*) nuclear agreements (IAEA, 2016) that came into force in January 2016. Verification measures are integrated as so-called **safeguards**. They enable IAEA staff members to get access to nuclear and research facilities, shut down and seal critical industrial hardware, install surveillance cameras, 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)[4], 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).

It is administered by the **Organisation 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 ratified by the treaty members in a dedicated Verification Annex.

This annex contains detailed explanations of verification measures, 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). Key elements 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 prohibited military "delivery systems"[5].

Regarding former technological developments that military forces had used, 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.

International security policies must handle a situation in which military forces are quickly adopting and considering cyberspace the next military domain where defensive

[4] 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".

[5] Tucker (Tucker, 1998) gives a comprehensive overview.

and offensive measures are necessary. More and more military forces are establishing dedicated cyber commands (UNIDIR, 2013), and alliances are fostering the establishment of collective capacities for military engagements. For example, NATO decided in 2016 that cyberspace is an essential domain that needs to be covered by 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 that takes into account the specific features of cyberspace. Although some suggestions have been made, such as the work of the ***Tallinn Manual*** (Schmitt, 2013 and Schmitt, 2017) 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. The most far-reaching step in this regard took place within the framework of the consensus report of the UN Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security, in which the general validity of binding rules of international law was also established in cyberspace.

This situation is tense on the one hand because 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 their networking hardware as critical infrastructure (for example, see EU, 2008) (see Chapter 14 "*Resilient Critical Infrastructures*" and 15 "*Security of Critical Information Infrastructures*"). Concerning 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, its military application to a great extent. 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 potentially destructive powers and to establish their own military programmes to counter this situation and keep pace.

11.2 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 former lessons learned and apply them to the current situation. Unfortunately, this approach soon reveals that cyberspace has 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.

11.2.1 The Problems of Counting Data in a Virtual, Distributed Space

Cyberspace is, by design, a "virtual" domain that abstracts a space from a specific actual 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 data is, on the one hand, theoretically stored and processed by a specific IT system, which usually has a geographical location and falls under a specific national legislation. On the other hand, especially in the 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[6] 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 of the number of objects. Digital data does not produce any 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 12 "*Attribution of Cyber Attacks*") that – in a nutshell – describes the problems and the ambiguity of assigning any activity within cyberspace to its origin and the presumed actor that intentionally performed this activity[7].

11.2.2 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*

[6] 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.

[7] 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).

Information Technology: Research, Development and Governance"). The term describes the feature of specific goods[8] 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 detail because only their precise usage decides whether it affects negotiated agreements. Popular examples of dual-use goods are biological agents or other essential materials for vaccines necessary for civilian health-care reasons and medical research, but they can also be used for military purposes. Defining lists of such goods and their necessary special verification into agreements has been performed for several decades for nuclear, chemical and biological goods. Its most famous 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 establishing national export controls. The agreement was extended in 2013 to cover so-called **intrusion software**, which is "specially designed or modified to avoid detection by 'monitoring tools', or to defeat 'protective countermeasures', of a computer or network capable device" (Wassenaar, 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 the number of specific materials, the necessary equipment or specific military delivery systems that can be controlled. This is impossible for cyberspace because both the hard- and software and their extent are the same for civil, economic and military purposes.

11.2.3 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 defence- and those that primarily serve offensive purposes. Such differentiation could be employed for regulating and verifying the trade, possession and usage of respective cyber capacities. Nevertheless, as pointed out before, IT goods have no obvious distinction due to their dual-use character. Even dedicated offensive tools like malware or software exploits are necessary to test and

[8] 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.

increase the cyber security of one's own IT systems. Popular examples for this case are so-called **penetration testing tools**, i.e. software specifically designed to attack and penetrate IT systems and networks to detect flaws, weaknesses and security problems. These tools are important instruments for IT security practitioners, and their regulation can affect the protection of IT systems. Their detection during potential inspections does not 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, 2024).

11.2.4 Established Verification Measures and Their Problems When Applied to Cyberspace

When considering cyberspace's technical characteristics, the previous glimpse at established verification measures of other technological developments already predicts that applying or projecting these measures directly will certainly not work for this new domain (Pawlak, 2016). Nevertheless, to understand how practicably 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 compliance with agreements, and although the previous examples illustrated that they strongly differ between various kinds of situations, all of them contain some of the following four restrictions and principles (Neuneck, 2012):

Geographical restrictions that regulate the allowed or prohibited location of specific items are checked by locating and visually monitoring them (this might include ultraviolet and x-ray imaging as well as aerial and satellite photography).

1. Limitations in terms of the overall number or even the complete prohibition of the possession of items are verified by counting and cataloguing them
2. 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
3. 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 virtuality of cyberspace undermines the principles of geographical restrictions. Even if the hardware itself always has a physical representation, data storage and processing cannot be reasonably attributed to a geographical location. Also, where hardware can be monitored and controlled, it is not the hardware but the software and its usage that differs between legitimate or a (theoretically) prohibited 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** (Hargreaves & Chivers, 2010), 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 rendered nearly impossible by its dual-use character. It is practically impossible 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 unlawful behaviour need to implement measures of checking the specific application of IT goods, which is not practically implementable.

One principle in which cyberspace mainly 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 commonplace but renders any approaches of limiting specific items useless. The strong dual-use character again interferes with this regulation approach for devices like IT hardware that theoretically can be counted.[9]

The principle that seems most suitable to be projected to cyberspace is the definition of any thresholds as part of verification regimes. This paradigmatically builds on the idea

[9] 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.

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.

11.3 Approaches to Verification for Cyberspace

Despite the problems that have been pointed out in the previous sections, verification for cyberspace has one substantial advantage over other domains. In contrast to air, space, sea and land, cyberspace is an entirely 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 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.

11.3.1 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”[10] within the IT infrastructure the measure can be performed and to what extent it needs any hardware or software alteration. Regarding 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.

[10] 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.

The first set of measurable parameters applies to the extent of the hardware of IT systems and networks and are, compared to later discussed usage-centric monitoring, quite rough. Instead, they represent the overall size of a facility, are physically apparent, hard to disguise or manipulate and visible for monitoring. They qualify for roughly estimating the storage or processing capacities, monitoring the tendency of technological developments of facilities as well as revealing the establishment of new cyber capacities or similar significant changes.[11] 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

Besides this list, other characteristics like the CPU, network processing power as well as available storage capacities could be used as parameters. But as already pointed out, these are more difficult to gather because measuring these values requires direct monitoring personnel access 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. Therefore, these parameters qualify for the real-time control of cyber operations and activities. In terms of necessary infrastructure adjustments, these parameters can also be gathered from outside by extending existing infrastructures without 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
2. The usage of anonymisation services
3. The usage of exploits for known security problems of IT devices and software

[11] As an example, the analysts of the so-called Mandiant report (FireEye, 2013) monitored among other parameters the extension of network bandwidth capacities and the necessary infrastructures in Beijing. They used their observations to harden their conclusion, that the Chinese army hosts one at least one cyber unit in Beijing, the so-called PLA unit 61,398, which is suspected to have been the cyber attacker behind many incidents against US companies, in this area.

11.3.2 Approaches for Verification Measures in Cyberspace

The previous section showed that IT systems provide measurable parameters that can be used to develop and establish monitoring procedures. Three important aspects that affect their technical applicability and the potential political acceptance of these measures by treaty parties need to be considered for their deployment. 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 particularly fertile for the cyberspace domain due to the already discussed dual-use character of its technologies, where the long history of IT security research often 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 tamper-proof storage of the logged data that will be discussed later. As pointed out, values of these parameters need to be collected and stored over a relevant time because their primary explanatory power lies in indicating significant infrastructure changes.

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 lightweight version can gather so-called **metadata** of outbound and inbound network connections. This metadata is information 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 they are 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 often already implemented for IT security or law-enforcement reasons.[12] 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, and 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.[13]

An important strategy of many cyber operations is their secrecy. 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 routing 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 so that the endpoint of the connection does not have any information about its origin. Anonymisation technologies effectively undermine the approach of linking cyber operations to their origin and, therefore, provide a possibility to avoid verification measures. 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.[14]

[12] 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.

[13] 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.

[14] Tor is designed to blend in with regular data traffic and look like normal HTTPS connections. On the other hand, tools that track network traffic and 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 (2010).

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 is an essential 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 – mainly 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; therefore, the usage of unknown vulnerabilities – so-called **zero-day exploits** – cannot be monitored directly. Furthermore, exploits and the delivered payloads can be encoded in a way that common detection mechanisms are rendered ineffective.

Nonetheless, verification often happens based on stored logged information collected over a specific period and analysed later. Even though recent studies show that zero-day exploits often stay undetected for several years[15], this provides at least an approach to put the activities of actors under observation. It must also be regarded from the perspective that, as stated before, most 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, 2024).

11.3.3 Implementation of Verification Measures

An important question regarding the described current state of cyberspace verification measures is whether 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 critical task since the early days of commercial applications of IT systems. Therefore, many technological developments have been established that can be used, and it is merely a question of how the results of these monitoring measures are interpreted. Where IT security aims to detect unwanted intru-

[15] See the RAND study (Ablon & Bogart, 2017) as an example. The study calculated an average life span of 6.9 years for zero-day exploits. This is put into perspective by other key findings of the study that "only 25 percent of vulnerabilities do not survive to 1.51 years, and only 25 percent live more than 9.5 years [and that for] a given stockpile of zero-day vulnerabilities, after a year, approximately 5.7 percent have been publicly discovered and disclosed by another entity".

sions or malicious activities that try to infiltrate a network from the outside, verification measures detect prohibited 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 adopting these technologies for verification is the validity of the logged information and its tamper-proof storage. This kind of technical verification for streams of logging data is a concept that has already been described as an "audit log" or "audit trail" for use cases in safety or security-critical scenarios (Schneier & Kelsey, 1998). More recent developments like the blockchain[16] are using cryptographic signatures and a so-called digital ledger, where each new data entry in the stream of logged information is verified by a digital key that is created based on the previous entries and then used to cryptographically sign the new entry (Putz et al., 2019). 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 in 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 significantly increases the necessary processing and storage capacities.

11.4 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 of arms control treaties and regimes that enable members or an authorised institution to check each other's compliance and guarantees the treaties' effectiveness. While verification as a tool 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

[16] A brief overview of digital and cryptographic signatures is given in "Introduction to Digital Signatures: The process & validity behind Digital Signature technology" (SecuredSigning, 2022).

specific adjustments of IT infrastructures because they are mostly already installed for IT security reasons.

- This provides a favourable position for both the establishment of the 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 effective the monitoring of specific variables is, mainly due to the fact that some discussed measurable parameters are mere generic values.
- About 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[17] to reflect its security- and stability-building intent.
- Finally, further research is also necessary to answer how measures can be developed or strengthened to prevent the circumvention or manipulation of monitoring.

11.5 Exercises

Exercise 11-1: Point out the specific features of cyberspace that hinder the application of established verification measures from former technologies.

Exercise 11-2: Explain which technical features and parameters of cyberspace that are practically measurable could be used for verification in cyberspace?

Exercise 11-3: Following the idea of a peace- and security-driven adaptation of cyberspace, which approaches of verification in cyberspace could be used, and what principles of this domain need to be changed for its application? Discuss and justify.

Exercise 11-4: Reflect on other approaches for verification in cyberspace that could be developed, and what are their technical preconditions would be.

Exercise 11-5: Assess the limitations and pitfalls of the presented verification approaches.

Exercise 11-6: Explain: How can the dual-use aspect of IT be resolved to differentiate between civilian and military usage of specific goods?

References

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.

[17] 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.

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. Double-day, New York, NY, USA.

Bibliography

Ablon, L., & Bogart, A. (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, E. (2007). *The Case for Deep Packet Inspection*. IT Business Edge.

Bazin, A. (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.

Boehme, P. (2008). *The Verification Regime of the Chemical Weapons Convention*. OPCW.

Bradner, S. (1999). Internet Engineering Task Force. In *Open Sources: Voices from the Open Source Revolution* (p. 280). O'Reilly & Associates.

EU, 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, pp.75–82.

European Parliament and Council of the European Union. (2006). *Directive 2006/24/EC*.

Granerud, A. O. (2010). *Identifying TLS abnormalities in Tor*. Gjøvik University College.

Guerrero-Saade, J. A., & Raiu, C. (2017). Walking in your enemy's shadow: When fourth-party collection becomes attribution hell. *Virus Bulletin Conference*.

Hargreaves, C., & Chivers, H. (2010). Detecting Hidden Encrypted Volumes. In B. De Decker & I. Schaumüller-Bichl (Eds.), *Communications and Multimedia Security* (Vol. 6109, pp. 233–244). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-13241-4_21

Hinck, G. (2018). *Wassenaar Export Controls on Surveillance Tools: New Exemptions for Vulnerability Research*. https://www.lawfaremedia.org/article/wassenaar-export-controls-surveillance-tools-new-exemptions-vulnerability-research

IAEA. (1961). *The agencys safeguards*. International Atomic Energy Agency. https://www.iaea.org/publications/factsheets/iaea-safeguards-overview

IAEA. (2016). *Iran and the IAEA: verification and monitoring under the JCPOA*. International Atomic Energy Agency.

Keohane, R. O., & Martin, L. L. (1995). The Promise of Institutionalist Theory. *International Security*, *20*(1), 39. https://doi.org/10.2307/2539214

Krasner, S. D. (Ed.). (1983). *International Regimes*. Cornell University Press.

Neuneck, G. (2017). 60 Jahre nuklearer—Prometheus oder Sisyphos? *Vereinte Nationen Magazin*.

Neuneck, G. (2012). Confidence Building Measures—Application to the Cyber Domain. *Cyber Security Conference*.

Pawlak, P. (2016). *Confidence-Building Measures in Cyberspace: Current Debates and Trends* (A.-M. Osula & H. Rõgias, Eds.; pp. 129–153). NATO CCD COE Publications.

Pimenta Rodrigues, G., de Oliveira Albuquerque, R., Gomes de Deus, F., de Sousa Jr., R., de Oliveira Júnior, G., García Villalba, L., & Kim, T.-H. (2017). Cybersecurity and Network Forensics: Analysis of Malicious Traffic towards a Honeynet with Deep Packet Inspection. *Applied Sciences*, *7*(10), 1082. https://doi.org/10.3390/app7101082

Putz, B., Menges, F., & Pernul, G. (2019). A secure and auditable logging infrastructure based on a permissioned blockchain. *Computers & Security*, *87*, 101602. https://doi.org/10.1016/j.cose.2019.101602

Schneier, B. (1996). *Applied Cryptography—Protocols, Algorithms, and Source Code in C*. John Wiley & Sons.

Schneier, B., & Kelsey, J. (1998). Cryptographic Support for Secure Logs on Untrusted Machines. *Proceedings of the 7th Conference on USENIX Security Symposium - Volume 7*, 4.

Secured Signing. (2022). Introduction to Digital Signatures: The Process & Validity behind Digital Signature Technology. https://www.securedsigning.com/blog/introduction-to-digital-signatures/.

Sherry, L., & Internet Task Force. (1996). Supporting a networked community of learners. *TechTrends*, *41*(4), 28–32.

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

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

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.

Tucker, J. B. (1998). Verification Provisions of the Chemical Weapons Convention and Their Relevance to the Biological Weapons Convention Biological Weapons Proliferation. Reasons for Concern, Courses of Action. *Stimson Center Report*, *24*.

UN, 1969. Vienna Convention on the Law of Treaties. *United Nations Treaty Series, 1155*, p. 331.

UNIDIR. (2013). *The Cyber Index—International Security Trends and Realities*. http://www.unidir.org/files/publications/pdfs/cyber-index-2013-en-463.pdf.

Verizon, 2024. 2024 Data Breach Investigations Report. [online] Verizon Enterprise. https://enterprise.verizon.com/resources/reports/dbir/ [Accessed 19 September 2024].

Wehberg, H. (1959). Pacta Sunt Servanda. *The American Journal of International Law*, *53*(4), 775. https://doi.org/10.2307/2195750