2016 8th International Conference on Cyber Conflict
Cyber Power
N.Pissanidis, H.Rõigas, M.Veenendaal (Eds.)
2016 © NATO CCD COE Publications, Tallinn

Arms Control in Cyberspace – Architecture for a Trust-Based Implementation Framework Based on Conventional Arms Control Methods

Markus Maybaum
Fraunhofer FKIE
Bonn, Germany
NATO CCD COE
Tallinn, Estonia
markus.maybaum@fkie.fraunhofer.de

Jens Tölle
Fraunhofer FKIE
Bonn, Germany
jens.toelle@fkie.fraunhofer.de

Abstract: This paper explores verification mechanisms, as well as confidence and security building measures, within the scope of existing conventional and strategic arms control regimes. In particular, it analyses the concepts of the Conventional Forces in Europe Treaty and the Vienna Document as an implementation regime for confidence and security building measures as well as the Open Skies Treaty, representing three major conventional arms control regimes. As an example for strategic arms control, we analyse the Chemical Weapons Convention. The objective of this paper is to identify those means and methods from these successful frameworks that can be adapted and potentially incorporated into a cyber-domain arms control regime. Based on this discussion, the authors suggest a general technical architecture for a trust-based future framework for arms control for cyberspace.

Keywords: *arms control, cyberspace, trusted computing, advanced trusted environment*

1. INTRODUCTION

Arms Control (AC) has been a success story since the late 1980s. While the first multilateral AC regimes focused more on conventional weapons and confidence-building measures in the air, space, and sea domains, recent treaties have been grander in scope. Consequently, arms control in cyberspace (ACCS) is seen as a necessary next step in building confidence between arising cyber powers operating within the cyber domain. AC is a commonly recognised instrument of security policy to avoid a competitive build-up of weapons between powers – an arms race [1]. Such a race is always costly for all sides. AC treaties have been negotiated and set into force to serve the purpose of limiting weapons stockpiles to a level that promises deterrence while conserving the economic and social resources of a state for other uses. When, more than 40 years ago, the Helsinki Final Act was signed, no one could really foresee the positive de-escalation between the main parties of the Cold War: NATO and the Warsaw Pact.

The story of success can be seen in more than three dozen AC treaties that have since been negotiated. However, recent developments in global security policy indicate a substantial change. Traditional AC has become subject to criticism and, especially from an American perspective, conventional AC in Europe has been disparaged [2]. This is not only due to the changed security situation in today's Europe but it is also an outcome of the inability of the international community to further advance these important security-enhancing instruments. The practice of hybrid warfare [3] has been underlying conventional AC regimes long before the Ukraine crisis – without any further action being taken or there being a need to implement suitable security mechanisms as a reaction to those demonstrated scenarios. Threats arising from cyber campaigns are no longer science fiction, and cyber weapons are undoubtedly a new instrument in conflict scenarios.

A. Aim

Besides obvious political obstacles, the development of an ACCS regime needs to cope with substantial practical and technical challenges. From the technical perspective, both security and privacy will be the most significant issues that must be addressed. On the practical side, the key question is enforcement: how can states make sure that any provisions are implemented? [4] Being aware of the complexity a regime to effectively monitor treaty compliance would require [5], an overwhelming majority of researchers in that field have all but given up; some have even concluded that ACCS in practice will be almost impossible [6].

This paper does not share that opinion. We think that, especially within the scope of building future cyberspace, there are several options to plan and implement instruments for ACCS. Conventional AC was seen as unlikely in the early 1980s, but less than a decade later the first international agreements on AC and verification had been put in place. Consequently, we see ACCS not only as a possibility, but also as a necessary next step in building confidence between rising cyber powers. It is also apparent that global players are starting to share our views to a certain extent. For example, we have seen international agreements such as the Wassenaar Arrangement [7] prohibiting the export of dual-use goods and technology, including computers and means of information security. A number of bilateral agreements have also been signed

between different nations. The first impulse towards an ACCS regime came from Russia, and their effort is still on-going [8]. In 2015 China and the USA negotiated a political agreement committing each country not to be the first to use cyber weapons to attack the other's critical infrastructure during peacetime [9], following the spirit of the Helsinki Final Act.

This paper's objective is to identify those means and methods that can be adapted from the successful AC frameworks and potentially incorporated into an ACCS regime. It recognises the functional gap between the identified and implemented requirements and evaluates the implications that arise from this difference analysis of future ACCS. As a proof of concept, a general technical architecture for a trust-based future framework for ACCS is suggested.

B. Definitions

Before we start analysing the AC regimes, we need to specify the key terms used in this paper. We will mostly refer to existing definitions that have been accepted in the academic community with some specific adaptations that may be required. For the purpose of this paper:

(a) A 'Weapon' is 'a means of warfare used in combat operations, including a gun, missile, bomb or other munitions, that is capable of causing either (i) injury to, or death of, persons; or (ii) damage to, or destruction of, objects' [10].
(b) A 'Weapon of Mass Destruction' is a nuclear weapon, a biological weapon or a chemical weapon as defined in [11].
(c) A 'Conventional Weapon' is a Weapon that is not a Weapon of Mass Destruction.
(d) A 'Cyber Weapon' (CyW) is a Weapon that would comprise any computer equipment or device that is designed, intended or used to have violent consequences; that is, to cause death or injury to persons or damage or destruction of objects [12].
(e) A 'Cyber Arm' as a CyW that is used in a computer network attack.

2. ANALYSING TRADITIONAL ARMS CONTROL REGIMES VS. CYBER ARMS CONTROL

In this chapter, we analyse well-known and established AC regimes. In particular, we identify ideas, methods, and techniques that have been deemed successful. We will explain the core objectives of four active AC regimes and highlight the main parameters based on which the treaty was designed and implemented. We will summarise these objectives and demonstrate the analogies to cyberspace.

A. The Conventional Forces in Europe Treaty and the Vienna Document

1) Main objectives

The Conventional Forces in Europe (CFE) Treaty, signed in November 1990, outlines provisions aimed at establishing a military balance between NATO and the Warsaw Pact, at a lower level of armaments [13]. It was negotiated during the late 1980s when NATO and the Warsaw Pact were both focusing on ending the arms race between East and West. The Vienna

Document (VD) is another regime of AC and confidence-building negotiated between the 57 participating states of the Organisation for Security and Co-operation in Europe (OSCE) [14]. It was adopted in 1990, together with the drafting of the CFE Treaty, and underwent its most recent fundamental revision in 2011 [15].

Both CFE and VD focus on conventional weapons in Europe. With CFE, the term Treaty Limited Equipment (TLE) was introduced with a broad scope, defining exact categories of weapons systems for which the treaty was supposed to be applicable. One of the typical characteristics of the CFE and other conventional AC treaties is that the main focus was laid on the weapons systems themselves and not, for example, on their ammunition. As we will see later in the context of the Chemical Weapons Convention (CWC), ammunition is only discussed separately when its damage potential is very high, such as for landmines or weapons of mass destruction. Based on the defined TLE, the CFE Treaty foresees an initial declaration of all TLE owned by a nation, its home location, and the military unit it is assigned to, including a full layout of force structure, followed by a yearly update. The establishment of a military balance between NATO and the Warsaw Pact was achieved by structural and geographical provisions such as defining limits for specific types of TLE, not only in total but also in certain concentric zones bearing in mind that for a successful attack TLE would have to be moved. Thus, the CFE regime in particular has an early warning function. To meet these limitations, equipment had to be destroyed or converted to non-military purposes.

The VD offers transparency and confidence-building by a declaration and inspection regime as well as mechanisms for peaceful conflict resolution. In signing the VD, the OSCE member states committed themselves to submit detailed information on their armed forces and principal weapons systems, their military budgets, their defence and force planning, and military exercises to the other state parties on an annual basis. OSCE states can conduct confidence-building inspections to verify this submitted information for compliance with the provisions of the VD. In addition to on-site inspection agreed in the CFE Treaty, VD allows so called ‘inspections of specified areas’ within the territory of the inspected state. This mechanism allows tracking of military activities that are taking place in these areas. The inspection team is entitled to check such an area on the ground and from the air. In addition, the VD foresees the invitation of observers to manoeuvres once they exceed a defined size, and it requires its member states to announce and present new weapons systems that they introduce.

The confidence-building function of both treaties was based on two main pillars: verification of both the declaration and yearly updates by on-site inspections, and social interaction between the inspection teams during their routine work. The CFE regime also established a so-called consultation group that had to deal with treaty interpretations and complaints. The consultation group was the official communication platform between the member nations. With Russia’s suspension of the CFE Treaty in 2007 [16] and subsequent loss of transparency around conventional forces, the politically binding procedures and related reports associated with the VD have become more important [17]. The signing parties agreed to hold an Annual Implementation Assessment Meeting where states are given a platform to discuss questions of implementation, operations, and questions that may have arisen from information that has

been exchanged or from findings during inspections. This meeting is hosted by the Forum for Security Cooperation [18] and has especially been used to discuss the situation in Ukraine.

2) Applying CFE and VD techniques in cyberspace

When thinking about applying CFE and VD techniques in cyberspace, the first issue that needs to be discussed is the definition of TLE and their categorisation. The core functions and principles of CFE refer to conventional weapons systems, counting these systems, and declaring possession and location of them. Considering our definition of a CyW and also other common definitions (e.g. 'software, firmware or hardware designed or applied to cause damage through the cyber domain' [19]), these core functions would only be applicable if hardware is involved; CyWs consisting of software are obviously uncountable [20].

The possession of such CyWs can be declared, of course, but any structural or geographical provisions or limits would not make sense or could not be subject to any form of on-site inspection. Thus, the early warning function, similar to the CFE does would not work efficiently in cyberspace and must be seen as a functional gap. The same applies to verification and social interaction; since any on-site inspection regime in combination with a declaration regime does not make sense within the scope of cyberspace, this function must also be seen as quite limited. The role of a consultation group within the scope of ACCS would be different from the role of a CFE consultation group; we will elaborate on this in more detail in section 3.

Nevertheless, the idea to allow areas to be subject to inspection can be of interest when thinking of inspecting parts of a state's cyberspace; for example, autonomous systems or parts of the national network infrastructure. Invitations to manoeuvres may also be an option when considering exercises or manoeuvres taking place in cyberspace, with the option of having observers present (physically or virtually). The challenge here is the level of detail an observer would be invited to see: would a cyber inspection team be granted read-access to inspection networks? How can espionage be prevented in such a setup? This raises essential questions that need to be answered.

B. Open Skies Treaty

1) Main objectives

The Open Skies (OS) Treaty established a regime of unarmed observation flights over the territories of states [21]. The idea of an airborne verification regime was born during the Cold War, but it never left the blueprint-stage due to mistrust within NATO and the Warsaw Pact; both sides were afraid of potential espionage. The OS Treaty was signed in March 1992. The ratification process in the Russian Federation took 10 years, as many technical details had to be discussed, tested and agreed on due to concerns over espionage. Finally, the treaty was put in force on January 1st, 2002.

The OS Treaty defines observation flights that an inspecting state party can conduct over the territory of another state party. The inspected state must provide airfields for that purpose from which those observation flights can be launched. One of the major objectives of OS is

the territorial scope: the observation flights can cover the entire country – thus, except from force majeure or natural conditions that would make flights impossible, no areas of a state party's territory can be excluded. An observation flight has to be notified three days in advance, specifying a point of entry (a predefined OS airfield). The detailed route is not submitted in advance, but negotiated with the inspected state party after arrival.

Besides quotas for observation flights and the notification of points of entry, the technical details and inspection of sensors were seen as critical during the negotiation of the OS Treaty. All parties involved were interested that the sensors of the observation aircraft could only be used for their dedicated purpose. This implies that all sensors must have specifically defined and technically assured parameters such as a maximum resolution which makes it possible to identify TLE, but not to record details of intelligence value. For this, all sensor configurations of an observation aircraft have to be certified, which in terms of the OS Treaty means that they have to be validated against a calibration target that allows an exact calculation of the camera resolution before an observation aircraft is permitted to conduct observation missions. The aircraft is also subject to inspection before missions to that ensure no additional sensors are hidden on board.

The confidence-building function for the OS Treaty is, by design, based mostly on social interaction between the mission teams during their routine work as well as the specialists working together during the specification of the technical parameters, calibration, and testing. Since the OS Treaty has no own declaration regime, it works more as a service for other verification regimes, supporting their confidence-building functions by providing the results of the observation missions to the teams. As an additional confidence-building measure, the results of the observation missions are provided to all OS Treaty participants.

2) Applying the OS Treaty's techniques in cyberspace

Referring to the main principles of the OS Treaty, some useful lessons can be learned for a possible ACCS regime. The first characteristic is the territorial scope: the entire country is subject to inspection. In the context of cyberspace, this would mean that the entire cyber infrastructure of a country would be subject to inspection. Still, software can be easily shipped and stored outside the country's borders, so there is still a functional gap as long as there are states not participating in such an ACCS regime. The same applies to small mass production hardware components such as microprocessors or other microchips. Malicious code can be stored and hidden easily, unless such a hidden functionality can reliably be found by technical means.

Serious challenges are posed if we take into account the technical details and certified sensors needed for ACCS mechanisms. Simply the overwhelming amount of known existing malware and the exponential increase of new examples may serve as an indicator that the traditional ways of finding malicious software by technical means have reached their limits. The functional gap we see here is a technical solution to reliably identify a CyW in cyberspace. For this, a new technology is required which experts can negotiate in order to find suitable technical parameters which can serve as the technical core of a future AC regime for cyberspace. Many

experts see this as the main argument against ACCS since finding reliable metrics for CyW detection is considered to be impossible, and large anti-virus companies have already given up this arms race [22]. So, the lesson we can learn from OS is that it could be used as an example to develop a cooperative approach of world leading experts working together to solve the malware problem. This would also surely support the confidence-building function we have seen in OS, gaining trust at the expert level by working together and developing a common platform that can make ACCS possible.

C. The Chemical Weapons Convention

1) Main objectives

The CWC is a strategic AC regime with a global scope [23] that was signed in 1993 and came into effect in 1997. It aims to eliminate all existing chemical weapons (CW) globally by prohibiting the development, production, acquisition, stockpiling, retention, transfer, or use of CWs by state parties. The first challenge with CWs is the exact definition of the TLE: what is a 'chemical weapon'?

The CWC's approach is to define lists of agents that fulfil certain requirements. In particular, typical agents for the use within the scope of chemical warfare are listed in relation to the Schedule 1 category. For the classification of CWs, the CWC introduced 3 categories [24]:

(a) Category 1: CWs based on Schedule 1 chemicals, including VX and Sarin (See below for an explanation of 'scheduled' chemicals);
(b) Category 2: CWs based on non-Schedule 1 chemicals, such as phosgene;
(c) Category 3: CWs including unfilled munitions, devices, and equipment designed specifically to employ CWs.

According to the CWC, a member state must declare its possession of CWs in an initial declaration. State parties are then obliged to plan and organise the destruction of their CWs. Since the destruction needs specific facilities to be built, and budget and administrative issues have to be solved, the joining state party negotiates an action plan with the Organisation for the Prohibition of Chemical Weapons (OPCW). The OPCW is an international implementation agency and has the role of a convention management entity supervising treaty implementation. In this action plan, a detailed time line with milestones is defined explaining how a member intends to eliminate its declared CWs. The progress of CWs' destruction is again subject to a notification regime, and the CWs, their storage facility, and the destruction facilities are subject to regular inspection.

Additionally, the CWC foresees so-called challenge inspections. If a CWC member accuses another member of false reporting of its CWs arsenal or any details of the negotiated action plan, the OPCW can initiate an area- or on-site inspection that the accused has to accept. Besides these technical specifications around a traditional declaration and inspection regime that have an obvious similarity to conventional AC systems, CWC has a global scope: so far, only four states (Egypt, Israel, North Korea, and South Sudan [25]) have not ratified the treaty.

This shows remarkable political will to do away with this class of weapon of mass destruction.

2) Applying CWC techniques in cyberspace

The core principle of CWC is not to focus on describing the CWs themselves, but to describe the agents these weapons carry as a payload. The three categories introduced in the CWC define a priority list. It is based on the severity of impact a weapon could have which is determined by the specific payload of the weapon.

CyWs can be seen as similar: the detailed malicious function may be undetectable until the weapon is launched, but the impact of a CyW in a specific attack scenario can be described in pre-defined metrics [26]. In general, the impact of a CyW will be either a breach of integrity of a system, or limitation to its availability; this is what an ACCS regime will have to detect.

The political will is also a major objective when thinking about regulating cyberspace [20]: will a majority of states be willing to share the idea of a peaceful use of the cyber domain? Will they be willing to invest in a common supervising entity such as the OPCW governing the future development of secure cyberspace and prohibiting warfare within the net? An action plan for weapon destruction will not be required if the common goal of nations is the peaceful cooperative use of modern information infrastructure.

The development of reliable technologies capable of monitoring malicious activities will be the key to success, so the question arising from the idea of the CWC relates back to a core technical problem: will we be able to develop a technical framework that allows us to identify CyWs in cyberspace? What would be the equivalent of a CWC challenge inspection? Would it be a reliable procedure based on globally trusted digital forensics that can prove or disprove a CyW engagement? We will address these questions in the following section.

3. CONCEPTS FOR THE IMPLEMENTATION OF A CYBER ARMS REGIME BASED ON TRADITIONAL ARMS CONTROL METHODS

When analysing the ideas and techniques of traditional AC in the context of cyberspace, we found potential analogies but also functional gaps that we see as requirements for a future ACCS regime. We discuss the analogies and gaps in the next section. Based on these findings, we then introduce a technical framework that we believe is helpful to make ACCS possible.

A. Analogies and gaps of traditional arms control vs. arms control in cyberspace

The core function of the first conventional AC treaties was the establishment of an early warning function, realising that the preparation of an armed attack in preparation of a future war or armed conflict would require movement and assembling of significant parts of a state's armed forces. In cyberspace, the situation is different.

CyWs cannot be clearly identified by an inspection team or a technical sensor due to their

characteristics. In the worst case, a CyW consists of numerous distributed pieces of information that are assembled at the moment of the attack. Thus, an approach to finding an early warning mechanism does not appear promising. Since, at the same time, a pre-planned cyber operation can be launched almost instantaneously, an ACCS regime would need a real-time warning mechanism. What makes it even worse is that without knowing potential patterns of CyWs, any information could be a potential suspect. Looking for 'dangerous parts' of software, similar to the list of dangerous agents within the CWC, is like searching for a needle in a hay stack, especially if encryption needs to be considered. Thus CyWs can only be detected if they have been seen before, or simply by coincidence. The dangerous weapons – based on the principle of unknown vulnerability exploitation – can mostly not be found this way.

On the other hand, hope cannot be a plan, meaning that for any further thought about an AC regime at the technical level, the focus at the current state of technology should not be on detecting the CyW itself but on its engagement or effect. In this respect, inspection of sites or areas, as we have seen in the traditional AC regimes, are unrealistic. The earliest possible time for a reliable detection of a CyW is when it is fully assembled and has a payload. This requires AC taking place at the 'speed of cyber'; we call this ad-hoc arms control. Technically speaking, we need to identify the breach of system integrity or the crippling of its availability. This can be achieved by technical means, as we will demonstrate in section 4.

The need to discover such integrity breaches or availability degrades already implies the use of sensors, as we have seen CWC and especially in OS before. Within the OPCW, experts of all nations share their knowledge on dangerous agents as well as on technologies of detection and CW disposal. The OS Treaty regime became successful precisely because technical experts of all parties worked together, and had the same goal of making it possible. This was achieved by openness, technical concepts, and a fully transparent framework. This also requires a technology that fulfils the aims and scope of AC which would not allow back doors or unauthorised use at the same time, as is the case with cyber espionage. In the context of cyberspace, we see the challenges arising from the development and implementation of such a framework as being even more complex. We think that an ACCS regime can be possible, if the political will is there and if persuading ideas for such a framework can be adequately promoted. One of the main challenges within this process will be the involvement of a majority of states around the globe, since geographical limitations within the scope of cyberspace are meaningless.

B. A common interest of states for stability and peace in cyberspace?

We can learn from the CWC that if mankind recognises an imminent threat beyond national borders, a common policy may be achieved. The objective of ACCS is to limit or stop a cyber arms race and to permit the peaceful use of cyberspace, which should be in the common interest of everybody. This is rather easy to demonstrate simply when showing people what a step back from the concepts of a digital information society to an analogue world would entail. Taking away the Internet from mankind would have a global impact on civilisation with unforeseeable consequences [27], providing reason enough to preserve this new territory and to establish common international security concepts and policies, including AC as an established concept of success. We see national legal frameworks and regulations, and authorities being assigned

responsibilities for national parts of the cyber domain. However, we do not see its equivalent for consultation boards as we have seen them for the traditional arms regimes. First discussions on confidence-building measures have been made at OSCE level [28], but so far we do not see groups of policy and technical experts cooperating to work on concepts for stability and peace in cyberspace. Cooperation between policy and technical experts is of major importance. The OS Treaty showed that establishing trust is the key, and that cooperation during the development of the technical framework was partly more effective in achieving confidence-building than the actual implementation of the jointly developed concepts.

By its nature, trust in cyberspace has a very technical dimension. The key is the development of reliable technologies. These technologies need to be transparent, so that their functionality can be understood by all parties involved. The technologies must guarantee the demands for confidentiality of states as well as their citizens and entities. The main challenge and obstacles with the Biological Warfare Convention (BWC) [29], for example, addressed exactly this point: the expert groups negotiating the implementation details were not able to find suitable technologies and procedures to establish an effective AC regime due to the difficulties with guaranteeing confidentiality of the business and state secrets of the inspected party. As a result, the BWC negotiations about the establishment of a binding verification regime have never succeeded and hopes for a global and verifiable prohibition of biological weapons, similar to the CWC, have unfortunately been abandoned. This will be another challenge for an ACCS regime: finding a technical solution to identify integrity breaches without getting to know too many details of the breached system. We will discuss and propose a possible solution addressing this concern in the next section.

Besides trust in technology, trust in science is a mandatory requirement when thinking about ACCS. Whereas in traditional arms control regimes the procedures and technologies could be demonstrated and made understandable to a broad audience, these concepts in cyberspace are significantly more abstract and a closer understanding cannot be expected either by political leaders or military decision makers. If the design of a future ACCS regime is too technically sophisticated, these decision making levels need to rely on experts they may or may not have. Taking this psychological aspect into account, we also see the requirement to design the technical framework based on technical standards that are internationally recognised and accepted. In our suggestions for a technical framework for future ACCS, we therefore refer to commonly developed international standards and enhance these standards to fulfil the necessary requirements.

4. A FUTURE ADVANCED TRUST ENVIRONMENT FOR ARMS CONTROL

Simply put, the core problem with a possible ACCS is to find the needle in the haystack without knowing what the needle looks like. In traditional AC, we know the needle, and in strategic regimes such as BWC and CWC we know at least the shape of the needle. Therefore, we can determine if certain identifiable parts can be part of such a needle. In the cyber domain, we do not know anything but the fact that we look for something having the function of a needle.

CyWs – like any other code – consists of many small 'puzzle' pieces of code. No one can say for sure if a piece of code is malicious without seeing the entire picture – or at least a big part of it.

All AC regimes with verification mechanisms we have seen so far work with a 'black list' describing the limited or prohibited item they control. This will not work with the infinite amount of potential malicious code pieces that can be merged to an uncountable number of different samples. In order to find working solutions, we have to reverse our thinking. A successful technical implementation framework for future ACCS will thus have to adopt a 'white list' approach that focuses on identifying good code from malicious code. This requires trust relationships in the levels of human-machine as well as machine-machine that have not yet been implemented in a broad scope.

As an enabler for future ACCS, we therefore suggest the development of a technical framework which we call Future Advanced Trusted Environment (FATE) to implement the requirements we defined in section 3.A based on technical white-listing. As we have already concluded, a peaceful use of cyberspace can be made possible by the early detection and prevention of the use of CyWs. Taking this into account, the application of CyWs in our context has to be technically understood as the execution of malicious code that breaks the integrity of the affected system or at least degrades its availability. Thus, we are here focusing on the challenges arising from this task.

To deliver FATE, both the computers as well as the communication links between the computers have to establish a Trusted Network (TN) that would technically implement such white-listing. The general idea behind this concept is that FATE must enable the real-time detection of any integrity breach during system operations as well as monitor and report any limitations on availability. Probing of availability and deriving a Common Availability Pictures (CAP) is daily practice in Computer Emergency Response Teams and Network Operation Centres around the globe. Monitoring integrity breaches at the tactical level and compiling a Common Integrity Picture (CIP) to form a Common Cyber Situation Picture (CCSP) is a functional gap that needs to be solved.

If the CIP is effectively established, good code can then be distinguished from malicious code (CyWs) even without knowing the details of the software products being used or the data being processed. This can be achieved by obtaining technical metadata that can be used to annotate programs and data. In current standard architectures, any code is processed. Malware is brought in as data and it hijacks a regular control flow by deviating the processing of the application into malicious code. Our approach is to counter the software exploitation exactly at that point: by defining which branches of an application control flow are legitimate during execution (white-listing), we can prohibit any control-flow deviation into malware – no matter what system is being used and what software is being executed.

Being aware of the need for such a white-listing-based architecture, we analysed the established concept of Trusted Computing (TC) [30] as the state-of-the-art hardware architecture in that field. In TC, all installed software (including BIOS and drivers) is digitally signed and any

unauthorised modifications can be recognised. Based on its hardware extension, the Trusted Platform Module (TPM), we have successfully demonstrated that the control flow integrity of processes running on Windows and Linux computers can be reliably monitored [31]. Common integrity breaches caused by CyWs can be identified and countered in real-time.

In the context of AC, the information on the integrity breach (potential CyW) needs to be gathered and collected in a CIP for which reliable trusted communication needs to be established. In particular, the problem of Man-In-The-Middle (MITM) attacks [32] as well as racing conditions between the TPM and the potential malware needs to be solved; we need to establish TNs of trusted machines following the white-listing principle. Peer instances communicating with the system and belonging to the same TN implementing this approach have to be reliably informed about the integrity breach within these TNs. Technically, with the concept of a trusted link-layer protocol establishing secure communication between TPM modules of the peering system, we were able to demonstrate MITM-resistant TNs that are specifically designed to win a racing condition against malware. More specifically, we demonstrated and proved the concept at the example of the Address Resolution Protocol (ARP) – Trusted-ARP (TARP) which we implemented in an extension module for the TPM – an Attack Recognition Module (ARM) [33]. This indicates that we are able to ensure CyW detection by applying the white-listing principle at process level within an entire network segment (and, in theory, with no limits in scalability).

What is needed to make this work? The technical management of such a TN has to be ensured by a local and trustworthy module in every computer that has a secured communication channel to both the communication modules used for exchange of information with peer computers as well as to the internal module detecting integrity breaches. Our suggested approach comprises of technical standard messages to manage peer membership as well as alert message distribution and acknowledgements to ensure reliability. Technically, and from an abstract view, the TN is a link layer network between ARM modules.

From an organisational point of view, FATE-based TNs can be built-up in a similar manner as state-of-the-art networks. A viable option would be to use local networks within organisations, mashed networks connecting these organisations, and, due to their scalability, set up TNs for entire nations. CIPs can then be generated at any granularity level (local network, Internet Service Providers at different TIER-levels, national and international governmental entities or organisations). A CCSP as a combination of CIPs and CAPs allows reliable recognition of CyWs without the need of knowing details of software or data being stored or processed on the affected system. We believe that such a CCSP can be a technical solution for a verification regime to make ACCS work.

5. SUMMARY, CONCLUSIONS AND WAY AHEAD

This paper presented a suggestion for the blue print of an architecture following the white-listing principle to support Arms Control in Cyberspace, based on the requirements derived from concepts and experiences from conventional arms control methods and treaties. A short

description of conventional arms control approaches was followed by a more detailed analysis of four treaties: The Conventional Forces in Europe Treaty (CFE); the Vienna Document (VD); the Open Skies Treaty (OS Treaty); and the Chemical Weapons Convention (CWC). All these agreements point out important aspects when discussing Arms Control in cyberspace, starting by the concept of Treaty Limited Equipment in CFE, followed by the concepts of transparency and confidence-building from VD, followed by the concept of considering the full territorial scope as well as the usage of certified sensors in OS, and finally the discrimination of carrier and payload in CWC.

Based on this analysis, one of the key questions of arms control in cyberspace is whether it is possible to implement a technical approach that allows us to reliably detect any engagement with Cyber Weapons (CyWs). We also showed that trust is a key requirement for any implementation approach. In its technical part, the paper therefore described a trust-based framework based on an extension of the Trusted Computing concept that reliably enables the monitoring and reporting of potential integrity breaches. Based on this mechanism, it is possible to generate situation pictures capable of seeing CyWs in sub-ordinated network structures. We still see research effort necessary within the scope of post-breach activity. Obviously, there are still numerous technical challenges we have to take into account. The points we see as most important to be further elaborated in the future are:

(a) How to react to alert messages/detected integrity violations.
(b) How to defend against Denial-of-Service (DoS) attacks.
(c) How to integrate this concept into operating systems and application processes in a user-friendly manner.

We also have to consider the practical obstacles. For example, in case of a potential treaty violation in CWC, a challenge inspection would be a suitable instrument to prove or disprove the presence of a chemical weapon. What would be the suitable follow-up activity for a detected integrity-breach? Recently we spent some research effort on digital forensics techniques, which we believe to be another very useful enhancement of the Trusted Computing technology and a useful instrument for arms control in cyberspace. As a possible solution, we can collect evidence of the CyW engagement, digitally sign it, and export it to an inspection-site for further investigation [33].

Of course, being familiar with the technical concepts of the Future Advanced Trusted Environment (FATE) we introduced in this paper, this mechanism can also be abused to launch DoS-attacks against FATE-protected systems. We discussed possible solutions for such kinds of scenarios in [34], but further research on this issue is required; we will elaborate on this in more detail in a separate publication. The biggest technical obstacle for successful implementation of our suggestion is the need to adapt operating systems to this new technology. We can implement the required enhancements to operating systems such as Linux since source codes are available, but what about the big proprietary operating systems? Will big vendors such as Microsoft or Apple be willing to support such a new standard?

Our main argument to further promote the idea of our FATE-architecture is the fact that we

can measure integrity without the need of knowing details of the system, either of the software installed or the data being processed. The confidentiality of system contents makes us believe that the necessary political will for an ACCS regime may be achieved. We feel that the concepts can be demonstrated to both the technical thought leaders and political decision makers around the globe and that the core ideas of our proposal will be understood. We also think that the system is not necessarily in conflict with other possible intentions of a state (e.g. active cyber operations), which in daily political business can also be a blocking point – depending on the priority of interests. We think that ACCS can be a next successful step in the history of arms control. To make it a story of success, we will continue developing the FATE framework by designing a software prototype as a proof-of-concept.

REFERENCES

[1] Paletta D., Yadron D. and Valentino-Devires J., Cyberwar Ignites a New Arms Race, in: Wall Street Journal – web portal, October 2015. Available: http://www.wsj.com/articles/cyberwarignitesanewarmsrace1444611128.

[2] Govan G. G., Conventional Arms Control in Europe: Some Thoughts About an Uncertain Future, in: Deep Cuts Issue Brief #5 – Conventional Arms Control in Europe, July 2015. Available: http://deepcuts.org/files/pdf/Deep_Cuts_Issue_Brief5_Conventional_Arms_Control_in_Europe%281%29.pdf.

[3] Van Puyvelde D., Hybrid war – does it even exist?, in: NATO Review magazine, 2015. Available: http://www.nato.int/docu/Review/2015/Also-in-2015/hybrid-modern-future-warfare-russia-ukraine/EN/index.htm.

[4] Denning D. E., Obstacles and Options for Cyber Arms Controls, presented at Arms Control in Cyberspace, Heinrich Böll Foundation, Berlin, Germany, June 2001. Available: http://faculty.nps.edu/dedennin/publications/berlin.pdf.

[5] Arimatsu L., A Treaty for Governing Cyber-Weapons: Potential Benefits and Practical Limitations, in: Proceedings of the 4th International Conference on Cyber Conflct, June 2012. Available: https://ccdcoe.org/cycon/2012/proceedings/d3r1s6_arimatsu.pdf.

[6] Nye J. S. Jr., The World Needs New Norms on Cyberwarfare, in: Washington Post – web portal, October 2015. Available: https://www.washingtonpost.com/opinions/the-world-needs-an-arms-control-treaty-for-cybersecurity/2015/10/01/20c3e970-66dd-11e5-9223-70cb36460919_story.html.

[7] The Wassenaar Arrangement on Export Controls for Conventional Arms and Dual-Use Goods and Technologies – web portal, 2016. Available: http://www.wassenaar.org/.

[8] Arquilla J., From Russia Without Love, in: Communications of the ACM, June 2015. Available: http://cacm.acm.org/blogs/blog-cacm/187854-from-russia-without-love/fulltext.

[9] Sanger D. E., U.S. and China Seek Arms Deal for Cyberspace, The New York Times, 19. September 2015. Available: http://www.nytimes.com/2015/09/20/world/asia/us-and-china-seek-arms-deal-for-cyberspace.html.

[10] Manual on International Law Applicable to Air and Missile Warfare, 2009. Available:http://ihlresearch.org/amw/HPCR%20Manual.pdf.

[11] Schneider B., Definition of 'Weapon of Mass Destruction', in: Schneider on Security, 2009. Available: https://www.schneier.com/blog/archives/2009/04/definition_of_w.html.

[12] Boothby W. H., Methods and Means of Cyber Warfare, in: International Law Studies – U.S. Naval War College, Vol 89, 2013. Available: https://www.hsdl.org/?view&did=734393.

[13] Organisation for Security and Co-operation in Europe, Treaty on Conventional Armed Forces in Europe, November 1990. Available: http://www.osce.org/library/14087.

[14] Organisation for Security and Co-operation in Europe, Vienna Document 2011, December 2011. Available: http://www.osce.org/fsc/86597.

[15] Organisation for Security and Co-operation in Europe, Agreement on Adaptation of the Treaty on Conventional Armed Forces in Europe, November 1999. Available: http://www.osce.org/library/14108.

[16] Russia Today, Russia completely ending activities under Conventional Armed Forces in Europe Treaty, March 2015. Available: https://www.rt.com/news/239409-russia-quits-conventional-europe/.

[17] Arms Control Association, Vienna Document 1999, in: Arms Control Association – web portal, August 2010. Available: https://www.armscontrol.org/factsheets/ViennaDoc99.

[18] Organisation for Security and Co-operation in Europe, Forum for Security Co-operation, in: OSCE web portal, January 2016. Available: http://www.osce.org/fsc.
[19] NATO Cooperative Cyber Defence Centre of Excellence, Cyber Definitions, January 2016. Available: https://ccdcoe.org/cyber-definitions.html.
[20] Geers K., Strategic Cyber Security, CCD COE Publications, 2011.
[21] Organisation for Security and Co-operation in Europe, Treaty on Open Skies, March 1992. Available: http://www.osce.org/library/141278.
[22] Gupta, A., The AntiVirus is Dead – says Symantec, The Windows Club Tech News – web portal, May 2014. Available: http://news.thewindowsclub.com/antivirus-dead-says-symantec-68180/.
[23] Organisation for the Prohibition of Chemical Weapons, Chemical Weapons Convention, April 1997. Available: https://www.opcw.org/chemical-weapons-convention/.
[24] Arms Control Association, The Chemical Weapons Convention (CWC) at a Glance, in: Arms Control Association – web portal, October 2015. Available: https://www.armscontrol.org/factsheets/cwcglance.
[25] Organisation for the Prohibition of Chemical Weapons, OPCW Member States, March 2016. Available: https://www.opcw.org/about-opcw/member-states/.
[26] Brangetto P., Caliscan E. and Maybaum, M., Responsive Cyber Defence, study & workshop presentation, presentation at the 6th International Conference on Cyber Conflict, CyCon 2014. Available: https://ccdcoe.org/cycon/2014/.
[27] Ziolkowski K., General Principles of International Law as Applicable in Cyberspace in Zilkowski, K., Peacetime Regime for State Activities in Cyberspace, NATO CCD COE Publications, 2013.
[28] Organisation for Security and Co-operation in Europe, Confidence-building measures to enhance cybersecurity in focus at OSCE meeting in Vienna, November 2014. Available: http://www.osce.org/cio/126475.
[29] United Nations, The Biological Weapons Convention, March 2016. Available: http://www.un.org/disarmament/WMD/Bio/.
[30] Trusted Computing Group, Trusted Computing - web portal, 2016. Available: http://www.trustedcomputinggroup.org/trusted_computing.
[31] Maybaum, M., Trusted Control Flow Integrity, in: Risiken kennen, Herausforderungen annehmen, Lösungen gestalten, SecuMedia Verlag, Gau-Algesheim, Germany, May 2015. (in German)
[32] IBM, Man-in-the-Middle, Trusteer web portal, 2016. Available: http://www.trusteer.com/en/glossary/man-in-the-middle-mitm.
[33] Maybaum, M. and Toelle, J., Trusted Forensics, to be published in: Proceedings to the 15th European Conference on Cyber Warfare and Security, ECCWS 2016, July 2016.
[34] Maybaum, M. and Toelle, J., ARMing the Trusted Platform Module, in: Proceedings to IEEE Military Communications Conference, MILCOM 2015, October 2015.