December 2018

Ethical Problems and Legal Issues in Development and Usage Autonomous Adversaries in Cyber Domain

Muhammad Mudassar YAMIN [1], and Basel KATT
Department of Information Security and Communication Technology, Norwegian University of Science and Technology, Norway

Abstract. An autonomous adversaries in cyber domain are new type of adversaries present in a cyber security exercise. Traditionally, adversaries in cyber security exercises are human who perform the roles of attackers and defenders. However, this is changing with time and autonomous adversaries are starting to appear in the cyber domain. The aim of this survey paper is to provide an overview of autonomous adversaries in cyber domain, furthermore ethical problems and legal issues related with the development and the usage of autonomous adversaries in cyber domain will be discussed.

Keywords. Autonomous adversaries, cyber domain, ethics problems, legal issues

1. Introduction

A cyber security exercise, is an exercise that is designed to evaluate the performance of cyber attackers an and defenders in a given scenario. The cyber security exercise life cycle contains planning, dry run, execution, evaluation and repetitions. We identified that the cyber security exercise life cycle is quite inefficient [1] [2] and found out that one way to remove these inefficiencies can be achieved by autonomous execution of adversaries present in a cyber security exercise. However, before the development of such autonomous adversaries we want to identify the type of ethical problems and legal issues posed by such research. In order to understand the ethical problems and legal issues associated with autonomous adversaries, first, we need to understand the types of autonomous systems. We identified three types [3] of autonomous system in the literature, details of which are given below:

- Supervised Autonomous Systems
  Autonomous systems that operate under the supervision of a human, the system is depended on human decision making for changing mission requirements. This system works in an *human in the loop* manner.

[1]Corresponding Author: Muhammad Mudassar Yamin, Department of Information Security and Communication Technology, Norwegian University of Science and Technology, Gjovik 2815, Norway; E-mail: muhammad.m.yamin@ntnu.no

- Monitored Autonomous Systems
  Autonomous systems that operate independently, however there actions are being monitored by humans. Humans intervene when they find that the actions of the autonomous system are undesirable. This type of system works in a *human on the loop* manner.
- Fully Autonomous Systems
  Autonomous systems that operate independently without human interference. Their decision making is based upon predefined rule set up by humans. This type of system works in *human out of the loop* manner

We want to develop the autonomous cyber adversaries just for educational purposes, however, we understand the potential malicious capabilities the proposed technology offers. Therefore, in this work we conducted a brief survey in order to answer the following legal and ethical question posed by such research activity.

- To identify current status of autonomous cyber adversaries' usage and development.
- To identify the legal status in development and usage of autonomous cyber adversaries.
- To identify the ethical guidelines in development and usage of autonomous cyber adversaries.
- To identify civil liabilities and unintended consequences caused by development and usage of autonomous cyber adversaries.

We used *autonomous cyber weapon* as a synonym for *autonomous cyber adversary* as the literature offered more material on autonomous cyber weapon, however it should be noted that a weapon can be used for offence as well as defence, thus, creating an adversarial environment. The rest of the paper is structured as follows. We first share the related work related to autonomous weapon systems. Then, we highlight the methodology which we use for the brief survey. After that we discuss the current status of autonomous adversary and the ethical and legal issues identified in the literature. finally we analyze the identified findings to answer the above stated questions and conclude the article.

2. Related Work

During the survey no related work focusing on our research topic was identified. Yet we identified a multitude of survey articles related to ethical and legal issues and challenges in cyber warfare and autonomous unmanned vehicles which are closely related to our research topic. Robinson et al. [4] in 2015 discussed ethical and legal issues in usage of autonomous cyber weapons in cyber warfare from a technical prospective. According to him, automating cyber offensive and defensive capability is a likely possibility for nation states. He highlighted the research from Caton et al. [5] according to which *"automated cyber weapons remove human decision making and could turn a bad situation into a catastrophic one"*. However, the problems which will arise by autonomous cyber weapons need to be tackle by a multidisciplinary research efforts from different doamins like laws, ethics and computer science.

Schuller et al. [6] in 2015 discussed the development and usage of autonomous weapons' systems with respect to international humanitarian laws. According to the au-

thors *"Autonomous weapon systems (AWS) are the most militarily significant yet legally elusive challenge to international humanitarian law (IHL) since the proliferation of cyber operations"*. Authors argued that the artificial intelligence and learning capabilities of these systems are quite different compare to traditional weapon systems. This creates new legal requirements for governing the development and usage of such system. Authors further elaborated that the commonalities of traditional and autonomous weapon systems can be governed under the same laws, however, the principles upon which the law is applied on traditional and autonomous weapon system will be quite different.

Lucas et al. [7] in 2014 discussed the legal and ethical precepts of emerging military technologies. The emerging military technologies include autonomous weapons. The researchers answered eleven precepts on the development and usage of autonomous weapons systems.

- Mission Legality
  Conducting a mission with a autonomous weapon should be justifiable by international law.
- Unnecessary Risk
  If conducting mission with autonomous weapon reduces the risk to human life, then autonomous weapon should be preferred.
- Moral Asymmetry of Adversaries
  There is no law which dictates confrontation with a technologically equivalent adversary.
- Greatest Proportional Compliance
  Use of autonomous system should be complaint with international law in distinguishing between a combatant and non combatant.
- Arkin Test
  The autonomous system complements the human agent performing or replace the human agent
- Non-Delegation of Authority and Accountability
  The action of autonomous system should be instructed by human agent who is responsible for consequences
- Due Care
  The autonomous system should be designed and developed under justifiable ethical guide lines and moral principles
- Product Liability
  The developer of the autonomous system should be liable for unintended consequences of there product usages
- Criminal Negligence
  The autonomous system should be designed to overcome the effects of negligence in operations
- Bench-marking
  The autonomous system should be tested against a set of defined standards.
- Orientation and Legal Compliance
  The human agent interacting with autonomous agent should know the safety, legal and compliance issues related to its operations.

researchers concluded that in case of a conflict where usage of traditional weapons are justifiable by law, the justification will extend to not just traditional weapons but cyber

weapons as well. In comparison to traditional weapons, cyber weapons provides a non lethal force enhancement capability which limits collateral damage. If the results of both cyber weapon and traditional weapons are equivalent the the usage of cyber weapons in obligatory.

3. Methodology

To identify relevant literature for the survey, keyword based search is used. We started with ”autonomous cyber weapons”, ”autonomous cyber weapons” with ”ethics” and ”autonomous cyber weapons” with ”laws”. We searched these keywords in academic databases like IEEE and ACM [8] to identify initial literature for the survey. After the initial collection of literature, we realized that the topics fall into interdisciplinary filed of study. Therefore we used Google scholar to relevant literature for relevant disciplines like defence and law. Although we identified many relevant material during the research we only used indexed research articles. We performed the collection of literature with keyword based search, however, repeating the same process will not yield the same results [9]. Therefore we are including the inclusion and exclusion criteria of research articles, to reduce the variation of results in literature searching.

3.1. Inclusion Criteria

We set the following inclusion criteria to for the survey:

- Articles written in English
- Articles directly related to autonomous cyber weapons.
- Articles that addresses ethical or legal, issues and challenges in usage of autonomous cyber weapons.

3.2. Exclusion Criteria

Due to large amount of identified literature we set the following exclusion criteria:

- Articles that mentions autonomous cyber weapons but are not directly related to them.
- Conference abstracts, book reviews, conference info, discussion, editorials, mini reviews, news and short communications.

3.3. Quality of Articles

The collected articles were evaluated against our defined criteria five quality assurance matrices. Points are allocated on a scale of one to five where five is considered as the highest value. The articles which scored the most on our defined criteria are given priority. The criteria which we used is given below:

- Reputation of publication channel, publication channels which are well known and recognized by academia scored higher in our criteria.
- Citation of article, articles with more citations given higher score in our criteria.
- Relevance of article content related to survey topic

- Publication date of articles, recently published articles received higher score compare to older articles.
- Number of references used to build the arguments in the article, articles scored higher with more references

4. Autonomous Cyber Adversaries

In this section we present the unclassified autonomous cyber adversaries that we identified in the literature. We identified two types of autonomous adversaries, one performs offensive cyber security operation, while the other performs defensive cyber security operation. Details of both are given below:

4.1. Offensive Autonomous Cyber Adversaries

- SC2RAM [10]
  SC2RAM (Simulated Cognitive Cyber Red-team Attacker Model) is developed to mimic the red team execution steps in a cyber security-exercise. It combines the cognitive capabilities of humans with the efficiency and reproduce-ability of a computer program. It can perform DoS (Denial-of-Service) attack on a given network. It is still at prototyping stage and is deployed at Michigan cyber-range and being used for testing and training purposes.
- SVED [11]
  SVED (Scanning, Vulnerabilities, Exploits and Detection) utilizes freely available exploit tools such as metasploit and nmap and automate their operations to execute red team activities autonomously in a cyber security exercise. SVED is deployed at CRATE cyber range and used for testing and training purposes.
- Stuxnet [12]
  Stuxnet was the most advance cyber weapon ever used in a military engagement. It is a piece of malware, that once injected in the target autonomously spread and compromise the whole network. It was designed to physically destroy the assigned military target by exploiting vulnerability in cyber physical infrastructure.

4.2. Defensive Autonomous Cyber Adversaries

- Intelligent Autonomous Agents for Cyber Defense [13]
  Researchers at United State army research laboratory proposed a model of an intelligent autonomous cyber defender. According to the proposed model the agent should be able to understand the environment in which it is operating. The agent should be able to identify malicious and non malicious changes in the environment and should autonomously act upon malicious change. The agent should be able to manage trust relationship with other operating agents and humans through a communication medium. Finally, the agent need to asses its performance and make necessary changes in its operating behavior to improve the performance for achieving the set objectives.
- Immuno-Inspired Autonomic System for Cyber Defense [14]
  A human immune system inspired automatic cyber defense model, that change systems security settings based upon evolving threats. The system is conceptu-

alized to be fully autonomous in nature and able to with stand current and new cyber attacks by adapting to the security challenges posed by evolving threats.

- VIAssist [15]
  A cyber security defender assistant that collects and visualizes data related to cyber security events. It automates the cognitive function required to analyze cyber security event data and presents the processed information to human defender to make final decision on the cyber security event.

Autonomous cyber adversaries are similar autonomous physical weapons, however they operates on the junction of cyber physical space. The action in cyber space have consequences in physical space, therefore we consider the existing ethical problem and legal issues for autonomous weapons systems to map it with autonomous cyber adversaries.

5. Ethical Problems

What is ethically acceptable for the usage of autonomous cyber adversaries? Which principals govern the morality in development of autonomous cyber adversaries? We analyzed ongoing research to answer these questions. We first describe the ethical foundation of using autonomous weapons systems. After that we analyze the military impact of such systems. Finally we focus on there effects of cyber operations.

International organization and research institutes have put significant efforts in determining the ethical foundation of development and usage of autonomous weapon systems. One such effort is led by joint efforts of Human Rights Watch (HRW), International Human Rights Council (IHRC), and Harvard Law School. Since November 2012, these efforts produced three detailed reports regarding the threats presented by "autonomous weapons systems". They characterize it as completely autonomous weapons with *human out of the loop* that "have the capacity to identify and engage with their target without significant human intervention. Their first report, *Losing Humanity: The Case against Killer Robots*(November 2012) [16] forms a case for the likelihood of autonomous weapon systems getting to be reality within 20 to 30 years also as an earnestness for their suggestion for suitable national and worldwide measures to forbid the advancement, development, and usage of completely autonomous weapons. The second report *Shaking the Foundations, The Human Rights Implications of Killer Robots*(May 2014) [17] stresses the moral ideas of human dignity and respect of human life. There report concludes that, fully autonomous weapon systems would not respect human life and dignity. The failure to maintain this basic guideline of human rights brings up genuine concern about the possibility of enabling a autonomous weapons to take a human life. The third and latest report, *Mind the Gap: The Lack of Accountability for Killer Robots* [18] (April 2015) concludes that the developers and operators of autonomous weapons system that breaks the laws should be held accountable for their actions.

The 2014 *NATO Allied Command Transformation* report on utilization of autonomous weapon system [5] discussed the moral issues of using autonomous weapon in the military. They concluded that the use of autonomous weapons system provides a unique opportunity to reduce risk to human life by not putting a human being in line of fire. It boost the moral of the fighting force by providing disposable alternative resources to conduct risky operations. Furthermore, autonomous weapons systems doesn't suffers from the psychological stress caused by critical military tasks. Although some of these

weapons do not kill directly humans, they can attack infrastructures that in a longer term can, indirectly, cause great human damage and sufferings.

Stuxnet [12] is the perfect example of a autonomous cyber weapon usage. Stuxnet was designed to destroy the uranium enrichment centrifuges of Iranian nuclear plants. Achieving this feat physically would have required deployment military personals in hostile territory. This could put human life in danger. However, with Stuxnet this feat is achieved without firing a single bullet. Stuxnet was the first known usage of a cyber physical weapons, it performed its intended mission, however, during its execution it also affected multiple civilian targets which raise the requirements for ethical conduct of such weapons under international law which. Similarly when cyber weapons are leaked they were used by adversaries and cyber criminals for monitory gains [21].

6. Legal Issues

What is legally acceptable for the usage of autonomous cyber adversaries? Which laws governs the development of autonomous cyber adversaries? We analyzed ongoing research to answers these questions. We first analyzed the international laws available to govern the development of autonomous cyber adversaries. Secondly we analyzed United nations assessment on proliferation and development of autonomous cyber adversaries. Finally, we focus on the human factor in usage and control of autonomous adversaries.

Autonomous weapons creates a moral asymmetry between the two adversaries and according to Lucas [7] *there is no requirement of fairness or technological equality in carrying out justified international armed conflict or lawful domestic security operations*. He argued that International Security Assistance Force (ISAF) is fighting a vastly technologically inferior adversary in Afghanistan, which is acceptable by international law. The technological superiority provide distinct advantages over the potential adversaries which resulted in proliferation advance weapons capabilities. This led to development of autonomous weapons for physical and cyber space.

United Nation member states that signed Convention on Certain Conventional Weapons (CCW) a treaty for eliminating munitions that are considered as inhuman, are working on to assessment of the threats and dangers posed by autonomous weapons systems [19]. In their most recent Group of Governmental Experts (GGE) meeting in Aug 2018 Geneva, all members agreed that humans should always retain the control of autonomous weapon systems. However they failed to reach an agreement for the achievement of this objective. They agreed on further discussion on this topic in next GGE meeting in 2019.

Autonomous decision of engagement with a target will pose significant legal concerns. In SC2RAM [10] the cognitive decision making ability of an attacker is mapped in to a computer program, which makes independent decision in achieving the target set by a human. The implications of those decisions need to be considered before the usage of such autonomous weapon. This technology is currently being used only for testing and training purposes however this technology has the potential to be used in actual cyber engagements if required. Red Cross in there 2016 report on autonomous weapons [20] discussed the autonomy of weapons systems in lethal engagement of targets. According to Red Cross such technology is not deployed yet in theater of war but the technology is certainly not out of the reach.

7. Analysis

7.1. Current Status

Details of current status of autonomous cyber adversaries are shared in section 4. Currently, offensive autonomous cyber adversaries are being rapidly developed, and used for training, testing and exploitation purposes. On the other hand, most of the research for defensive autonomous adversaries is at designing stage. Most of the current research is funded by milliliters around the world for the achievement of technological edge over their human adversaries

7.2. Legal Status

Development and usage of autonomous cyber adversaries are currently in a legal gray area. There is currently no law exists that explicitly prohibits the development and usage of such technologies [19]. International efforts are being carried out to reach a consensus on regularizing the proliferation of such weapons, however, no common ground on this issue is achieved yet [19] [4].

7.3. Ethical Guideline

Autonomous cyber weapons provide unique opportunity in reducing human suffering by achieving the desired military objective in a non lethal way [5]. However, such weapons decrease the threshold of a conflict between adversaries [13]. Therefore, the usages should be governed by ethical and moral guidelines where commonalities with other weapons systems exists [6]. Human dignity and respect of human life should be given up-most priority while using such technologies [17].

7.4. Civil Liabilities and Unintended Consequences

Autonomous weapons system developers are responsible for unintended consequences of the autonomous system functionality and the autonomous weapon system user is responsible for criminal negligence [7]. The consensus of international community established that the decision making of a autonomous system should always be governed by human [19].

8. Conclusion

From the above analysis it can be concluded that there is no legal restriction in development of autonomous adversaries. However, the developers of such systems are liable for any intended consequence of such technologies. Similarly user are of such technologies are liable for any kind of criminal negligence in there operation. International consensus is already established that any kind of fully autonomous weapon system with *human out of the loop* is not acceptable. However, a philosophical question is still open for further research: *Do we wants to develop autonomous machine slaves that blindly follows human orders or do we want to develop autonomous machine that question the ethics and morality of human orders ?*

References

[1] Vykopal, Jan, et al. "Lessons learned from complex hands-on defence exercises in a cyber range." Frontiers in Education Conference (FIE). IEEE, 2017.

[2] Uckan Frnman, B., Koraeus, M., and Backman, S. (2015). The 2015 Report on National and International Cyber Security Exercises: Survey, Analysis and Recommendations.

[3] Anderson, J. M., Arbour, B., Arnold, R., Kadiofsky, T., Keeley, T., MacLeod, M. R., ... and Roorda, M. (2015). Autonomous Systems: Issues for Defence Policymakers. NATO SUPREME ALLIED COMMAND TRANSFORMATION NORFOLK VA NORFOLK.

[4] Robinson, M., Jones, K., and Janicke, H. (2015). Cyber warfare: Issues and challenges. Computers and security, 49, 70-94.

[5] Caton, J. L. (2013, June). Exploring the prudent limits of automated cyber attack. In Cyber conflict (CyCon), 2013 5th international conference on Cyber conflict (pp. 1-16). IEEE.

[6] Schuller, A. L. (2017). At the crossroads of control: The intersection of artificial intelligence in autonomous weapon systems with international humanitarian law. Harv. Nat'l Sec. J., 8, 379.

[7] Lucas Jr, G. R. (2014). Legal and ethical precepts governing emerging military technologies: Research and use. Amsterdam LF, 6, 23.

[8] Jesson, J., Matheson, L., and Lacey, F. M. (2011). Doing your literature review: Traditional and systematic techniques. Sage.

[9] Kitchenham, B., Brereton, P., Li, Z., Budgen, D., and Burn, A. (2011, April). Repeatability of systematic literature reviews. In Evaluation and Assessment in Software Engineering (EASE 2011), 15th Annual Conference on (pp. 46-55). IET.

[10] Jones, R. M., OGrady, R., Nicholson, D., Hoffman, R., Bunch, L., Bradshaw, J., and Bolton, A. (2015). Modeling and integrating cognitive agents within the emerging cyber domain. In Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) (Vol. 20).

[11] Holm, H., and Sommestad, T. (2016, November). Sved: Scanning, vulnerabilities, exploits and detection. In Military Communications Conference, MILCOM 2016-2016 IEEE (pp. 976-981). IEEE.

[12] Langner, R. (2011). Stuxnet: Dissecting a cyberwarfare weapon. IEEE Security and Privacy, 9(3), 49-51.

[13] Kott, A., Thomas, R., Draar, M., Kont, M., Poylisher, A., Blakely, B., ... and Rigaki, M. (2018). Toward Intelligent Autonomous Agents for Cyber Defense: Report of the 2017 Workshop by the North Atlantic Treaty Organization (NATO) Research Group IST-152-RTG. arXiv preprint arXiv:1804.07646.

[14] Dasgupta, D. (2007). Immuno-inspired autonomic system for cyber defense. information security technical report, 12(4), 235-241.

[15] Goodall, J. R., and Sowul, M. (2009, May). VIAssist: Visual analytics for cyber defense. In Technologies for Homeland Security, 2009. HST'09. IEEE Conference on (pp. 143-150). IEEE.

[16] Docherty, B. (2012). Losing humanity: The case against killer robots.

[17] Docherty, B. L. (2014). Shaking the Foundations: The Human Rights Implications of Killer Robots. Human Rights Watch.

[18] Docherty, B. L. (2015). Mind the gap: The lack of accountability for killer robots. Human Rights Watch.

[19] Bode, I., and Huelss, H. (2018). Autonomous weapons systems and changing norms in international relations. Review of International Studies, 1-21.

[20] Righetti, L., Pham, Q. C., Madhavan, R., and Chatila, R. (2018). Lethal Autonomous Weapon Systems [Ethical, Legal, and Societal Issues]. IEEE Robotics and Automation Magazine, 25(1), 123-126.

[21] Dwyer, A. C. "The NHS cyber-attack: A look at the complex environmental conditions of WannaCry." (2018): 25-26.