Recommendations Concerning the Selection of Artificial Intelligence Methods for Increasing of Cyber-Security

Roumen Trifonov[†]
Technical University
Sofia, Bulgaria
r_trifonov@tu-sofia.bg

Slavcho Manolov
Technical University
Sofia, Bulgaria
s_manolov@tu-sofia.bg

Georgi Tsochev
Technical University
Sofia, Bulgaria
gtsochev@tu-sofia.bg

Galya Pavlova
Technical University
Sofia, Bulgaria
raicheva@tu-sofia.bg

ABSTRACT

In the course of the implementation of a project funded by the Research Fund under the Ministry of Education and Science, on the base of multi-factor analysis of Cyber-security Threats and the adoption of military theory into Cyber-security procedures, the authors have formulated a new classification of the stages of the Cyber Defense and the classes of tasks that can be solved using Artificial Intelligence methods.

Based on theoretical models and experiments with specific prototypes, the authors make recommendations for the selection of Artificial Intelligence methods appropriate for the various phases of the Cyber Defense

CCS CONCEPTS

• Formal security models • Security requirements

KEYWORDS

Cyber Defense, Cyber Threat Intelligence, Artificial Intelligence methods

ACM Reference format:

Roumen Trifonov, Slavcho Manolov, Georgi Tsochev and Galya Pavlova. 2020. Recommendations Concerning the Selection of Artificial Intelligence Methods for Increasing of Cyber-Security. In *Proceedings of CompSysTech '20, June 19–20, 2020, Ruse, Bulgaria, 5 pages.* https://doi.org/10.1145/3407982.3407987

1 Introduction

The Faculty of Computer Systems and Technology of the Technical University - Sofia undertook in 2013 research in the field of application of Artificial Intelligence methods in Cyber-security. Since 2017, these studies have been supported by the Research Fund under the Ministry of and Science.

In the project Web site [1] as well as in the 14 publications and reports to authoritative international conferences mentioned in it, the constructed models, the experimental installations and the results of the experiments are described in detail. This article is intended to announce to the expert community a summary of these results, prepared as a set of recommendations. Due to the limitations of the conference format, they are given in a rather fragmentary form.

Beginning the work on this project, its contractors planned to experiment with the effective use of theoretically selected Artificial Intelligence methods to detect attacks, prevent intrusions, and more typical actions in Cyber Defense of Information Systems. Naturally, the research began with the assessment and classification of Cyber Threats [2], as well as more and more being put in the theory and practice of Cyber-security military concepts, such as "Cyber Threats Intelligence" [3] and the "Kill Chain" (Fig. 1) [4].

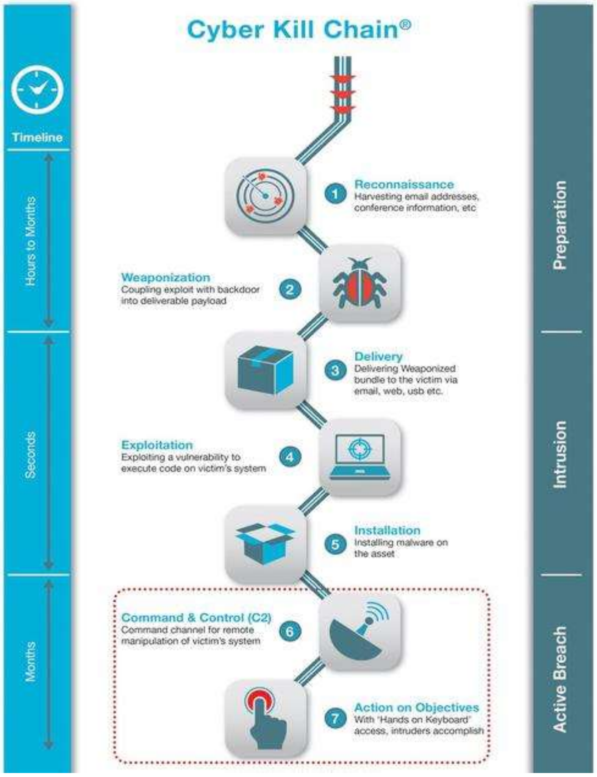
Cyber Kill Chain®
Timeline
Hours to Months
Seconds
Months
1 Reconnaissance
Harvesting email addresses, conference information, etc
2 Weaponization
Coupling exploit with backdoor into deliverable payload
3 Delivery
Delivering Weaponized bundle to the victim via email, web, usb etc.
4 Exploitation
Exploiting a vulnerability to execute code on victim's system
5 Installation
Installing malware on the asset
6 Command & Control (C2)
Command channel for remote manipulation of victim's system
7 Action on Objectives
With "Hands on Keyboard" access, intruders accomplish
Preparation
Intrusion
Active Breach

Figure 1: Cyber Kill Chain

CompSysTech '20, June 19–20, 2020, Ruse, Bulgaria

ACM ISBN 978-1-4503-7768-3/20/06…$15.00
https://doi.org/10.1145/3407982.3407986

But in the course of the research the project team came to three fundamental conclusions, which (within the referenced sources) have not been formulated so far in explicit form:

A. The Cyber Defense (depending on the circumstances and methods and tools used) can be divided into three main components:

A1. Immediate repulsion of attacks - in terms of the above mentioned military concepts adopted by the cybersecurity expert community, this can be referred to as so called Tactical Cyber Threats Intelligence. It is a process of monitoring events occurring in a computer system or network and analyzing signs of possible incidents that are violations or imminent threats of breaches of computer security policies, acceptable usage policies, or standard security practices;

A2. Predicting the actions of a possible opponent - this in the same terms can be referred to as so called Operational cyber intelligence, where its primary objective is to reduce the risk to the security of critical assets and activities of the organization by: defining the operational environment, describing the impact of the operational environment, identifying potential adversaries and identifying the intentions and actions of these potential adversaries. The Operational Cyber Intelligence is based on the Doctrine of Active Defense;

A3. Troubleshooting of consequences caused by Cyber-attacks - this can be referred to as so called Incident Handling. Globally, the concept of Computer Security Incident Management is widely accepted and implemented. It assists the systematic response to incidents (i.e. by following a consistent methodology for handling them) and helps staff minimize the loss or theft of information and the disruption of incident-related services. Another advantage of this concept is the ability to use the information received during incident processing to prepare for future incidents and provide greater protection for systems and data. As a formal procedure, incident handling is regulated by a number of international standards, such as: ITU Recommendations E.409 and H.1500, ISO 18044, and SR 800-61 of US National Institute for Standardization and technology (NIST).

B. International developments in the theory and practice of the use of Artificial Intelligence methods in the field of Cyber-security indicate that these methods are applicable at all stages of the cyber defense mentioned above. At the same time, the literature (and to a certain extent the research and experiments carried out within the scope of this project) clearly indicates that there is no universal method that is equally effective in all (or even most) possible applications.

The possibility (and the necessity) to select methods of Artificial Intelligence effective for a given class of tasks (a certain classification of these methods is introduced in the respective activity report of the project) requires, first of all, formulation of criteria for the selection of an effective method based on both literature sources and specific research and experiments.

C. In addition to the general conclusion described above about the motivation of the application of a particular Artificial Intelligence method to solve a specific Cyber Defense task, the team reached another fundamental conclusion within the project, which in a nutshell could be formulated as follows: the Artificial Intelligence method selected by certain criteria as the most effective for a particular task can be considered as basic. In most cases, its effectiveness could be enhanced by supplementing it with another appropriate method. The set of basic and complementary methods could be called a hybrid Artificial Intelligence method.

As mentioned above, this classification of the stages of cyber defense and the approaches of choosing specific methods of artificial intelligence for each of them are original contributions of the authors.

2 Evaluation of tasks solved through different Artificial Intelligence methods

According to a number of experts, one of the important areas of the fundamental and applied research in the field of Artificial Intelligence, and above all in its general theory, is the so-called "task oriented approach" to the research and development. The basic idea of this approach is that the whole activity of the subjects involved in the processes of application of Artificial Intelligence, it is appropriate to describe, model and design as a system of processes for solving various tasks. The effectiveness of this development is ultimately determined by what tasks and in what sequence they will be solved. Therefore, the qualitative and quantitative characteristics describing the tasks, as well as the means and ways of solving them, are of great importance for the creation of effective Artificial Intelligence systems.

This "task oriented approach" to the study of Artificial Intelligence systems is manifested in the definition of the types of solved problems, in the formulation of abstract means for solving them, as well as in the "task oriented" analysis of the various types of relationships between humans and Artificial Intelligence systems. In addition, one of the goals of this approach is to build structures capable of solving one or the other type of task, and another aim may be to determine the content of the training of users interacting with one or the other Artificial Intelligence system.

In this order of thought, for creation of systems for solving a particular type of problem it is of utmost importance to establish the convergence of the decision process.

The concrete assessment of the tasks in the field of Cyber Defense can be done after classifying these tasks according to two main criteria:

A. The nature of the action taken by the application of these methods:

A1. **Automated action** - this, as the name implies, is the operation of an Artificial Intelligence system without the direct human involvement and naturally, related to performance requirements, i.e. it applies above all in cases where response time is a critical element in Cyber Defense. This determines its application in the Tactical Cyber Intelligence (or immediate reflection of the attack).

To some extent, the automated action is also applicable in Incident Handling, but only partially (for example, when based on a formalized incident report (so-called Trouble Ticket or other standardized data structure) the system uniquely categorizes the incident as a result of one of the typed threats in the so-called "Treath Taxonomy" of ENISA [5].

A2. **Decision-making assistance** - in a large proportion of Cyber Defense Artificial Intelligence systems do not directly trigger action but are used to support Cyber-security-related decision-making (both fundamental decisions on infrastructure, security policies, etc., and one-off decisions on short-term actions).

B. The nature of the task - this classification is not as obvious as the above. It is an original contribution of the participants in the project - analyzing a large number of tasks solved using the methods of Artificial Intelligence, the authors concluded that they can be grouped into two large groups:

B1. **Anomalies detection** - detection of anomalies in the behavior of the communication network or its subjects, mutations in certain samples, etc. As practice shows, this type of task is solved mainly in the Tactical Cyber Intelligence;

B2. **Solving a classification problem** - in general, this involves either determining the affiliation of the object (situation, model, etc.) to one of the components of a commonly accepted classification scheme (or taxonomy), or identifying the object (situation, model, etc.).) by its characteristics compared to the characteristics of certain patterns. We consider that this type of task would be solved mainly in Incident Handling and, with great likelihood, in Operational Cyber Intelligence.

C. An additional feature - in distinguishing between the types of tasks (respectively, the selection of an effective Artificial Intelligence method), specific areas of application could be identified within each section depending on the structure of the threats and the agents of the threat.

3 Established criteria and concrete selection of effective methods for the various stages of Cyber Defense

It should be emphasized that according to the approved "Scientific Description" and "Work Program" of the project, its activities were entirely focused on Tactical Cyber Intelligence or direct detection and action to reflect attacks, to prevent unauthorized access to systems resources, to block malware, etc.

In order to carry out these actions, the project team has developed prototypes and laboratory productions, has realized experiments with simulation of real attacks, has analysed results and has undertook actions aimed at improving the work of the prototype.

Therefore, specific recommendations for the use of one or other Artificial Intelligence method apply only to this element of Cyber Defense (according to our classification).

The description and initial considerations of the other two stages of Cyber Defense formulated during the implementation of the project are beyond the scope of the current project and are presented for the purpose of outlining the problem.

3.1. For Tactical Cyber Intelligence

Traditionally, the processes of detection and prevention of Cyber-attacks are based on two basic principles: anomaly detection and abuse detection, although there is no significant difference in their characteristics. The intrusion detection mechanism typically detects the following suspicious actions I the network:

(a) attempts to use services blocked by firewalls;

b) unexpected requests, especially from unknown addresses;

c) unexpected encrypted messages;

d) overly active traffic from unknown servers and devices;

(e) significant changes from previous network operations;

(f) attempts to exploit known bugs or vulnerabilities;

(g) attempts to access of unknown users from unexpected addresses;

h) improper or suspicious use of administrative functions;

(i) significant changes in the user's usual activities, etc.

The detection of suspicious activity of users in network traffic is the most widespread applications in Machine Learning. The modern systems are more and more successful in detecting unusual events in large data streams and in solving standard analysis problems.

The type of attack detection depends on the nature of the threats used: "notify known", "known unknown" and "unknown unknown" [6]. The global expert community has formulated criteria for assessing the effectiveness of detection and the level of response. It is also crucial to strike the right balance between false positives and false negatives results. The false positives (so-called false alarms) can be no less damaging than false negatives ones.

The project team adopts the following approach of analyzing and comparing different methods of Artificial Intelligence: combining a basic criterion (or set of basic criteria) with additional criteria. For Tactical Cyber Intelligence the main criteria were selected as follows: maximum productivity (i.e. detection efficiency combined with performance level) and a minimum percentage of fals alarms.

We also have accepted additional criteria including: flexibility for use in different environments; generic methodology; the processing speed needed to analyze packet contents to exclude lost packets [7].

On the base of the multivariate analysis mentioned above, the authors selected Artificial Intelligence methods potentially suitable for these applications. The data obtained indicate that, when managing devices for detection and prevention of intrusions, the discovery of abnormal behavior, of attacks with unknown nature by a network of self-learning multi-agent systems shows higher efficiency and productivity than other methods. In addition, this method shows a minimum value of false-positive alarms. (Fig 2, 3).

The use of mobile agents may be further recommended, although their effect has not yet been as well studied as that of fixed agents. Mobile agents do not directly improve attack detection techniques, but they can reframe the techniques which are applied, thereby improving efficiency. Having agents visiting data warehouses and monitoring results is an ideal alternative that is appropriate for mobile agents' ability to minimize computation. In addition to reducing network load, the specialized agents can be focused on specific classes of intrusion, such as coordinated attacks, which occur over long periods of time from different sources.

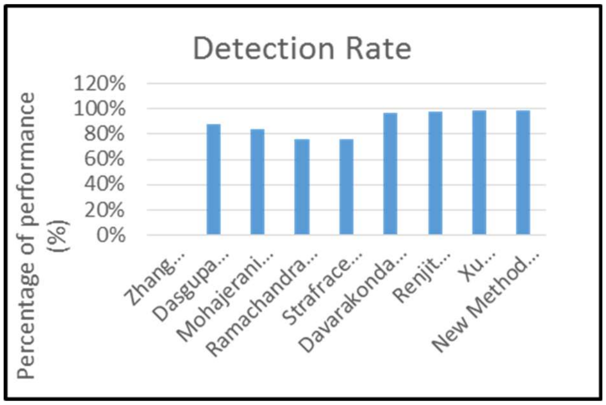

Figure 2: Detection Rate of various methods

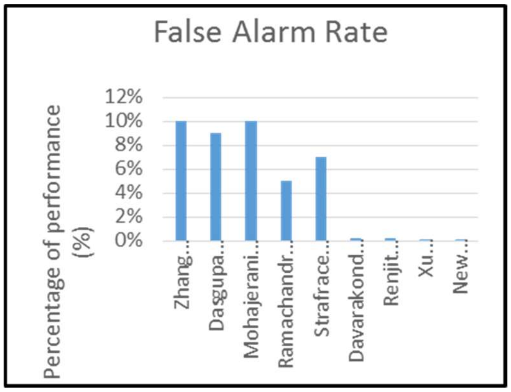

Figure 3: False Alarm Rate of various methods

In the experiments carried out in the course of the project, the basic method of a multi-agent system composed of self-learning stationary agents was supplemented by so called Reinforcement Learning (RL) and Fuzzy Sets (FS). A series of tests were conducted to determine whether the use of RL and FS could allow agents to learn to categorize normal and unusual activity in the network using information about the information flows. It turns out that in the case of an anomaly, this combination analyzes internally and updates the Q-value of the trainee agent realizing recursive iteration of its training cycle.

The experiments showed an increase in the detection accuracy of the hybrid model by an average of about 6% (Fig. 4).

3.2. For Operational Cyber Intelligence

Our hypothesis for the automation of Operational Cyber Intelligence is that this may be based on the behavioral model of the likely adversary based on his activity in the network and outbound traffic. Currently, in the absence of references for such studies, it is assumed that this behavior can be divided for the present into two basic types: hostile and non-hostile.

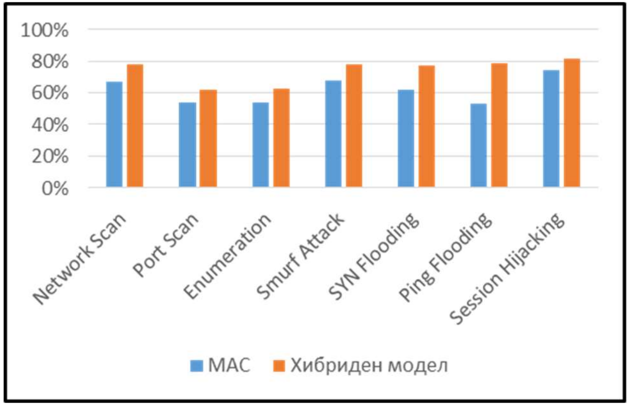

Figure 4: Detection Accuracy in case of various attacks

This supposes a closed-loop scenario consisting of four steps:

(a) measuring network activity using methods such as Packet Capture Appliances [8];

(b) pre-treatment;

(c) extracting specific characteristics called 'features';

(d) classification that specifies a class of feature set corresponding to the type of behavior identified.

On the base of literature survey, the Echo State Network (ESN) method (a class of recurrent neural networks) was proposed as a feature selection mechanism. The ESN structure consists of a reservoir of randomly coupled dynamic neurons with sigmoid nonlinearities (usually hyperbolic tangents) [9].

By examining the feasibility of training classifiers on a variety of topics, we suggest that the Sequential Feature Selection (SFS) procedure for behavior classification will be most appropriate. This can reduce the inherent variability of the data and can lead to high accuracy in recognizing the status of behavior between subjects [10].

3.3. For the Incident Handling phase

From the scarce literary sources describing attempts to apply Artificial Intelligence methods in Incident Handling the project team concluded that the main function of Artificial Intelligence currently can be focused to the solving of classification task, i.e. the unambiguous reference of the current incident to one of the elements of an approved classification scheme, whereby procedures and work processes are developed for each element of the scheme.

Obviously, the most important part of solving this classification task is finding so-called "features," i.e. characteristics that reflect adequately objective dependencies on classification status.

If we adhere to this approach, it should be noted that through the application of Artificial Intelligence in the Incident Handling must be solved both right and opposite task.

The right task can be summarized as follows: to find 'features' in the attributes contained in unified incident reports based on international standards (for example, the so-called Trouble Ticket [11]).

The reverse task seems more difficult - it involves changing the attributes of unified incident reports so that they more adequately reflect the affiliation of a particular incident to one element of the classification scheme. Here we can say that this task is not yet considered appropriate by the project team.

The analysis of literary sources has led us to recommend in future studies the "Reinforcement Learning" - a "trial and error" learning paradigm [12]. In cases where the model is unable to predict the value with a sufficiently high confidence, the result of this incident can be manually corrected by a person. In doing so, the relevant change event must be fixed and used to improve the model.

4. Conclusion

The recommendations for government, business and academia formulated in the final part of the project do not focus on the researched and experimented specific methods of Artificial Intelligence, but insist on the use of the above mentioned classification at the stage of Cyber-security and the approach to choosing appropriate methods of Artificial Intelligence on the base of clear criteria.

In addition, they stressed the need for some new areas of research, such as:

a) exploring the applications of Artificial Intelligence in Industrial Cyber-security, especially with a view to the EU initiative “Industry - 4.0”;

b) research for relationships between these ones applications in the various stages of Cyber-defense (for example, the ability of the application in the field of Operational Cyber Intelligence to support the operation of systems related to Tactical Cyber Intelligence).

ACKNOWLEDGMENTS

This research is realized and funded under the scientific-research project № H07/56 “Increasing the level of network and information security using intelligent methods” by the contract DH 07/4 with Bulgarian National Science Fund.

REFERENCES

[1] Roumen Trifonov, Ognian Nakov, Slavcho Manolov, Radoslav Yoshinov, Georgy Tsochev and Galya Pavlova. *2020,* Increasing the level of network and information security by using intelligent methods*. Research project funded by NSF, https://iti.tu-sofia.bg/en/national-science-funds-project-h-07-56/*

[2] Roumen. Trifonov, Slavcho Manolov, Georgy Tsochev, Galya Pavlova. 2018. New Approaches in the Examination of the Cyber Threats. Proceedings of the *International Conference on Information Technologies (Info-Tech 2018) 20-21 September 2018 Varna, Bulgaria*

[3] ENISA Threat Landscape Report 2016

[4] www.lockheedmartin.com/content/dam/lockheed/data/corporate/documents / LM-White-Paper-Intel-Driven-Defense.pdf

[5] ENISA Threat Taxonomy A Tool for structuring Threat Information January 2016

[6] National Cyber Security Strategy “Cyber Resilient Bulgaria” Sofia, 2016

[7] Roumen Trifonov, Slavcho Manolov,Radoslav Yoshinov, Georgy Tsochev, Galya Pavlova. 2017 An adequate response to new Cyber Security challenges through Artificial Intelligence methods. Application in Business and Economics, *WSEAS Transactions on Business and Economics, 14 pp. 272 – 281,E-ISSN: 2224-2889*

[8] Erik. Hjelmwik 2011. Passive Network Security Analysis with Network MinerForensic Focus, *http://www.forensicfocus.com/passive-network-security-analysis-networkminer*

[9] Mantas Lukosevicius and Herbert Jaeger 2009 Reservoir computing approaches torecurrent neural network training, *Computer Science Review, vol. 3, p.127-149, DOI: 10.1016/j.cosrev.2009.03.005*

[10] Isabelle Guyon and André 2003 Elisseeff;An Introduction to Variable and Feature Selection, *Journal of Machine LearningResearch, vol.3, DOI:10.1162/153244303322753616*

[11] Request for Comments 1297 ”Internal Integrated Trouble Ticket System – Functional Specification Wishlist”, Merit Network Inc., January 1992

[12] Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, Anil Anthony Bharath 2017. A Brief Survey of Deep Reinforcement Learning, *IEEE Signal Processing Magazine, Special Issue on Deep Learning for Image Understanding*, *DOI 10.1109/MSP.2017.2743240*