Artificial Intelligence in Cyber Threats Intelligence

Roumen Trifonov
Faculty Computer Systems and Technologies
Technical University of Sofia
Sofia, Bulgaria
r_trifonov@tu-sofia.bg

Ognyan Nakov
Faculty Computer Systems and Technologies
Technical University of Sofia
Sofia, Bulgaria
nakov@tu-sofia.bg

Valeri Mladenov
Faculty Automation
Technical University of Sofia
Sofia, Bulgaria
valerim@tu-sofia.bg

***Abstract*—In the field of Cyber Security there has been a transition from the stage of Cyber Criminality to the stage of Cyber War over the last few years. According to the new challenges, the expert community has two main approaches: to adopt the philosophy and methods of Military Intelligence, and to use Artificial Intelligence methods for counteraction of Cyber Attacks. This paper describes some of the results obtained at Technical University of Sofia in the implementation of project related to the application of intelligent methods for increasing the security in computer networks. The analysis of the feasibility of various Artificial Intelligence methods has shown that a method that is equally effective for all stages of the Cyber Intelligence cannot be identified. While for Tactical Cyber Threats Intelligence has been selected and experimented a Multi-Agent System, the Recurrent Neural Networks are offered for the needs of Operational Cyber Threats Intelligence.**

Keywords— Cyber Threats Intelligence, Artificial Intelligence, Behaviour Assessment, Neural Networks, Sequential Feature Selection, Remote Network Monitoring

I. INTRODUCTION

Over the last few years, the trends of transition of the Cyber Threats from the Cyber-Crime phase to the Cyber-War phase has also prompted an adequate transition of Cyber Defense techniques to military technology [1]. First of all, this concerns the perception in the analysis of the threats of the so-called "Cyber Kill Chain" model, as well as the application of traditional Military Intelligence Technology.

Furthermore, in the conditions where well-resourced and trained adversaries conduct multi-year intrusion campaigns targeting highly sensitive economic, proprietary, or national security information, the network defense techniques which leverage knowledge about these adversaries can create an intelligence feedback loop, enabling defenders to establish a state of information superiority which decreases the adversary's likelihood of success with each subsequent intrusion attempt. That's why according to the vast majority of experts, the qualitative transition to new Cyber Defense tools must involve the widespread use of Artificial Intelligence methods to analyze information exchanged, network flows, sources of threats, and to plan effective impact measures, including proactive ones.

Following these trends, the Faculty of Computer Systems and Technology at Technical University of Sofia began research on the application of intelligent methods for increasing the security in computer networks. An essential section of this investigation is dedicated to the Cyber Threat Intelligence. The present article summarizes some results of a research done by the project team.

II. BASIC FEATURES OF THE CYBER THREATS INTELLIGENCE PROBLEM FORMULATION

The Cyber Intelligence or, more precisely, Cyber Threats Intelligence (CTI) has the following definition in the draft Bulgarian National Cyber Security Strategy [2]:

- establishment of mechanisms and technical means to maintain an up-to-date picture of possible threats of different scale, sources and character, trends in geopolitical context development and relevant national cyber picture analysis and;
- development of capabilities to help identify attribution sources and take appropriate forms of protection and counteraction.

According to the documents of INSA (Intelligence and National Security Alliance) [3, 4, 5] the preparation of the intelligence in cyber operational environment is a systematic and continuous process of analyzing potential threats to detect a suspicious set of activities that may endanger systems, networks, information, employees, or customers by providing means to visualize and evaluate a number of specific penetration sensor inputs to bring up a particular threat. This process supports the organization's risk management strategy and decision-making in the area of information security. Its application identifies potential threats and assists security and risk managers selectively implement and maximize deep defense strategies by better understanding the critical points in time and space in the operating environment.

The Cyber Threats Intelligence Cycle [6] is a systematic, continuous process of analyzing potential threats to detect a suspicious set of activities that might threaten the organization's systems, networks, information, employees, or customers by providing a means of visualizing and assessing a number of specific intrusion sensor inputs and open source information to infer specific threat courses of action. The model supports the organization's risk management strategy and the information security group's decision-making. The application of the model identifies potential threat courses of action and helps the security and risk management leaders selectively apply and maximize a defence in depth strategy via a greater understanding of the organization's cyber threats at critical points in time and space in the operational environment by:

a) defining the operational environment;

b) describing the operational environment effects on network defense;

c) evaluating the cyber threats, and

d) developing cyber threat courses of action

978-1-5386-6477-3/18/$31.00 ©2018 IEEE European Union

Fig. 1 is a graphical representation of the Cyber Threat Intelligence Cycle.

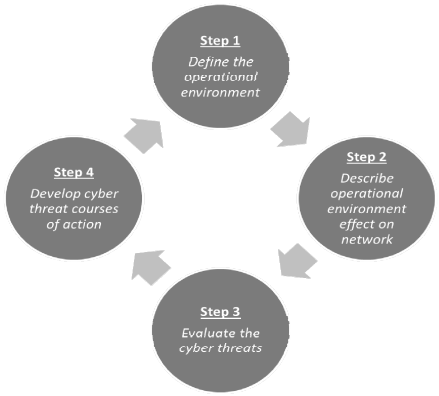

Fig. 1. Cyber Threat Intelligence Cycle

Like its military analogue, the Cyber Threats Intelligence is developed at three levels: strategic, operational, and tactical. For the purposes of this study, the second one is considered: INSA defines [5] the operational level as: "The level at which campaigns and major operations are planned, conducted, and sustained to achieve strategic objectives within theaters or other operational areas. At this level, actors build the capabilities needed to support the tactical operations. They maneuver in cyberspace to position capability where they need to in order to be effective in their tactical missions. At the operational level, an organization's operating environment can be described in terms of physical, logical, information and social layers".

III. METHODS OF ARTIFICIAL INTELLIGENCE IN NETWORK AND INFORMATION SECURITY

The essence of Artificial Intelligence (AI) is based on the statement that people's intelligence (the potential (inborn) ability of a conscious individual to conclude on a given information) can be described so precisely that it is machine-simulated. After several decades of research, AI is not only the subject of research or planning of some movement, but also of more complex and interdependent solutions. Artificial Intelligence is defined as the intelligence displayed by machines and / or software. This is an academic field of study exploring the goal of creating intelligence. The main issues explored by AI include reasoning, knowledge presentation, automated planning and scheduling, machine learning, natural language processing, computer vision, robotics and common intelligence.

AI enables us to develop autonomous computer solutions that adapt to their context of use, using self-management, self-tuning and self-configuration, self-diagnosis and self-healing. When it comes to the future of information security, AI looks like a very promising field of research that focuses on improving cyberspace security measures.

With rapid pace of development and the desire for more effective countermeasures, Artificial Intelligence comes as a natural solution to the problem of coping with the ever-growing number of network attacks. Applications in the field of AI are widely accepted by the modern information society. This interdisciplinary endeavor has created a joint link between computer specialists and network engineers in designing, simulating and developing network penetration patterns and their characteristics.

As mentioned in the introduction to this article, world practice has already noted a significant number of various Artificial Intelligence applications in computer security. Without trying for a comprehensive classification, we could divide these methods into two main directions:

A. Conditionally named "distributed" methods:

A1. Multi-Agent Systems of Intelligent Agents;

A2. Neural Networks;

A3. Artificial Immune Systems and Genetic Algorithms.

B. Conveniently named "compact" methods:

B1. Machine Learning Systems, including: associative methods, inductive logic programming, Bayes classification;

B2. Pattern recognition algorithms;

B3. Expert Systems;

B4. Fuzzy logic.

Having into account this variety of methods, it is of particular importance that adequate criteria are selected for the assessment and selection of a specific application for each specific solution. In the above mentioned project, the specification was carried out for two of the main sections of CTI. It should be noted here that within the project the application of Multi-Agent systems was chosen and experimented as the most appropriate method for the needs of the Tactical Cyber Intelligence.

IV. METHODS OF ARTIFICIAL INTELLIGENCE SUITABLE FOR OPERATIONAL CYBER THREATS INTELLIGENCE

The ultimate goal of Operational Cyber Intelligence is to reduce risk to an organization's critical mission and assets by: defining the operating environment; describing the impact of the operating environment; evaluating the adversary; and determining potential adversarial courses of action (COA). The Operational Cyber Intelligence provides a thread that links the probability and impact of a cyber attack with its strategic level implications by ensuring a coherent framework for analysis and prioritization of potential threats and vulnerabilities given the organization's threat environment. Operational Intelligence is based on the Doctrine of Active Defense. Instead of searching for information regarding a specific attack against the organization, it focuses on analyzing the opponents' combat doctrines, weapon systems and attack and operational scenarios. This approach shifts the center of gravity to the ability to respond and block the outcome of the attack within the organizational environment or in its immediate vicinity.

Our main idea was that the basis for the automation of the Operational CTI can be the behavioral model of the likely adversary. It should be emphasized that the problem of using artificial intelligence methods in the Operational CTI is a completely new matter, and systematized literary sources have not yet been found. Only, there are reports concerning the use of behavioral analysis based on machine learning by

the companies: Exabeam (USA), Darktrace (UK), CyberX (USA), Interset (Canada).

The TU-Sofia team concluded that the activity and the outgoing traffic in the network of the supposed adversary were to be the main source of information for building his behavioural model. This evokes analogies with the Non-Invasive Brain - Computer Interface whereby the physiological signals of the human brain (for example, through Electroencephalograms (EEGs)) can be used for human emotions evaluation [7].

Indeed, the streams of measured parameters received by n-number different IP addresses of the monitored object using RFC 1757 Remote Network Monitoring methods [8] can be compared to EEG with n-number of channels.

If this analogy is applied in practice, first of all, on the order of the classification model of emotions [9], a basic classification of the behavior of the possible adversary, based on the needs of our research, must be constructed,. 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.

In order to obtain the best possible performances, it is necessary to work with a smaller number of variables which describe some relevant properties of the data retrieved from the network. These variables are known as "features". Features can be aggregated into a vector known as "feature vector". Thus, feature extraction can be defined as an operation which transforms one or several signals into a feature vector. Identifying and extracting good features from signals is a crucial step, because otherwise the classification algorithm will have trouble identifying the class of these features, i.e., the behavioral state of the possible adversary. According to some researchers [10], it seems that the choice of a proper pre-processing and feature extraction method have more impact on the final performances than the selection of a good classification algorithm [16].

Therefore, following the analogy of the Brain-Computer Interface, two basic tasks have to be solved:

- to find a suitable approach to selecting characteristics from which to derive features suitable for behavioral interpretation and validation. In doing so, the necessary inter-subject discrimination of the features for the subsequent classification must be ensured;
- to build and optimize an ensemble of classifiers based on trained models to be used to assess behavior.

According to the researcher's scenario, design of the system of assessing the behaviour of the supposed adversary can consist of two main phases: 1) offline training phase to calibrate the system and 2) online phase which uses the system to recognize the type of behavior states and translate them into the computer commands. Both offline and online phases follow a closed-loop process, generally composed of six steps:

a) network activity measurement- this step consists in network surveillance of broadband Internet traffic (e-Mails, Web traffic, instant messengers, etc.) using methods, such Packet Capture Appliances Fig. 2 in order to obtain signals reflecting the opponent's intentions [11];

b) preprocessing - this step consists in cleaning and denoising input data to enhance the relevant information embedded in the signals;

c) feature extraction – this extraction aims at describing the signals by a few relevant values called "features";

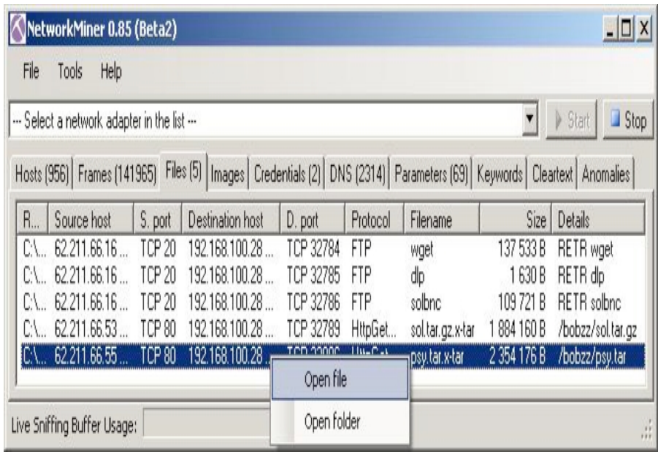

Fig. 2. Packet Capture Appliances

d) classification - this step assigns a class to a set of features extracted from the signals, which corresponds to the kind of behavioral state identified. This step can also be denoted as "feature translation". Classification algorithms are known as "classifiers" [15];

e) translation into a command/application - once the behavioral state is identified, a command is associated with this state in order to control a given application.

Once the data have been acquired, they are pre-processed to clean (de-noise) the signals and to enhance relevant information embedded in these signals. The pre-processing step aims at increasing the signal-to-noise ratio of the input signals.

To perform this pre-processing, various spatial-spectro-temporal filters [10] can be used. Naturally, numerous other pre-processing methods, which are more complex and more advanced, can be proposed and used. But in our initial experiments we were based on two of the most popular methods, namely, Independent Component Analysis (ICA) and Common Spatial Patterns (CSP) method.

Based on a study of literary sources, the Echo State Network (ESN) method was proposed as a mechanism for feature selection – this is a class of Recurrent Neural Networks where the so-called "Reservoir Computing" approach for training is formulated [12].

The basic structure of an ESN, presented in Fig. 3, consists of a reservoir of random connected dynamic neurons with sigmoid nonlinearities (usually hyperbolic tangent):

$$r(k) = f\,res(W\,inin(k) + W\,resr(k-1))$$

and a linear readout *f out* (usually identity function) at the output:

$$out\,(k) = f\,out\,(W\,out\,[in\,(k)\,r\,(k)])$$

Here *k* denotes discrete time instant; *in(k)* is a vector of network inputs, *r(k)* - a vector of the reservoir neurons states and *out(k)* – a vector of network outputs; *nin*, *nout* and *nr* are the dimensions of the corresponding vectors *in*, *out*

and *r* respectively; *W out* is a trainable *nout* (*nin*+*nr*) matrix; *W in* and *W res* are *nr nin* and *nr nr* matrices that are randomly generated and are not trainable. In some applications, direct connection from the input to the readout is omitted.

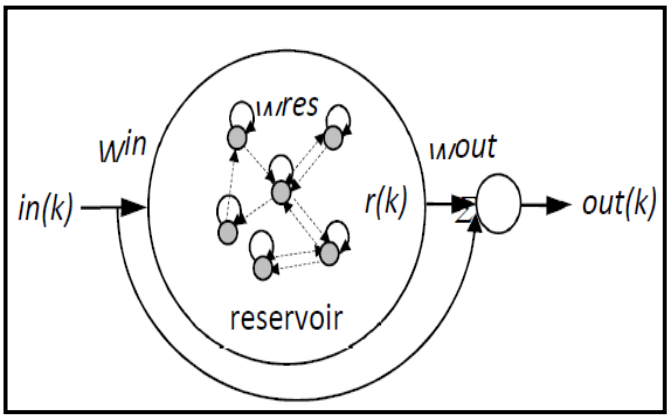

Fig. 3. Basic structure of an ESN

The main advantage of the ESN is the simplified training algorithm since only weights of the connections from the reservoir to the readout neurons are subject to training [13]. Thus instead of gradient descent learning much faster least squares method can be used.

We started on the presumption that using reservoir computing pre-training is beneficial for selecting the most relevant discriminative features and reaching state-of-the-art performance for subject independent recognition. The Reservoir Computing approach could be used not only for time series processing but also for high dimensional static data representation. Finally, the existing practice shows that IP-trained ESNs outperform pre-trained deep auto-encoders and can actually achieve almost 100% testing accuracy.

Exploring the feasibility of training cross-subject classifiers, we have settled on the Sequential Feature Selection (SFS) procedure [14] that reduces the inherent data variability and can lead to a high inter-subject behaviour status recognition accuracy. Starting from an empty set, SFS increments sequentially a new feature that best predicts the class at the current iteration. The process stops when there is no more improvement in the prediction. SFS is a very effective way to identify the dominant behavioral signatures across subjects. However, it is a computational heavy and time-consuming procedure, which was the main motivation to look for a computationally less intensive alternative.

The state of the art of the works described in this article can be defined as a transition from the development of a theoretical model to an experimental setting.

As the experiments are in their early stage, it is necessary to point out that the results are encouraging, but it is still too early to declare any definitive conclusions.

V. CONCLUSION

As can be seen from the above, the process of introducing Artificial Intelligence methods at the different levels of Cyber Threat Intelligence is at very different stages: while in Tactical Intelligence, it has long gone out of the phase of research and experiments and is used for building real effective systems, In the field of Operational Intelligence, these studies are in a very initial phase and require the commitment of substantial resources. Furthermore, the question arises as to the application of possible outcomes of Operational Iintelligence in the activity of Tactical Intelligence systems, which are intended to neutralize the immediate threats to computer systems and networks.

REFERENCES

[1] *ENISA Threats Landscape Report* 2016: 15 Top Cyber-Threats and Trends, ENISA, 2017

[2] Republic of Bulgaria: *National Cyber Security Strategy "Cyber Resilient Bulgaria 2020"*, 2016-03 NCSS Bulgaria final draft v 5 3, Bulgarian government, 2016

[3] *Cyber Intelligence: Setting the Landscape for an emerging Discipline, Intelligence and National Security Alliance* , INSA, 2011

[4] *Operational Level of Cyber Intelligence*, INSA, 2013

[5] *Operational Cyber Intelligence*, INSA, 2014

[6] Brian P. Kime *Threat Intelligence: Planning and Direction*, SANS Institute, 2015

[7] Liu Y., Sourina O. and Nguyen M. K., Real-time EEG-based human emotion recognition and visualization, in *Proceedings of the International Conference on Cyberworlds* (CW '10), 2010, Singapore

[8] [RFC 1757] Remote Network Monitoring Management Information Base, Carnegie Mellon University, 1995

[9] L. Bozhkov, P. Georgieva Classification models of emotional biosignals evoked while viewing affective pictures, in *Proceedings of the International Conference on Simulation and Modeling Methodologies, Technologies and Applications* (SIMULTECH), 2014, Vienna

[10] Hammon P.S. and Sa V.R. de, Preprocessing and meta-classification for brain-computer interfaces, *IEEE Transactions on Biomedical Engineering*, 54(3), 2007.

[11] Erik Hjelmvik, Passive Network Security Analysis with Network Miner, (IN)SECURE Magazine, no. 18, pp. 18–21, 2008.

[12] Lukosevicius M. and Jaeger H., Reservoir computing approaches to recurrent neural network training, *Computer Science Review*, vol. 3, 2009

[13] Koprinkova-Hristova P., Bozhkov L. and Georgieva P., Echo State Networks for feature selection in affective computing, in *Proceedings of the 13th International Conference on Practical Applications of Agents and Multi-Agent Systems (PAAMS)*, 3-5 June 2015, Spain

[14] Guyon I. and Elisseeff A. An Introduction to Variable and Feature Selection, *Journal of Machine Learning Research*, vol. 3, 2003.

[15] Hanan Hindy, David Brosset, Ethan Bayne, Amar Seeam, Christos Tachtatzis, Robert Atkinson, and XavierBellekens. 2018. A Taxonomy and Survey of Intrusion Detection System Design Techniques, Network Threatsand Datasets. 1, 1 (June 2018), 35 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

[16] Elike Hodo, Xavier Bellekens, Ephraim Iorkyase, Andrew Hamilton, Christos Tachtatzis and Robert Atkinson. Machine Learning Approach for Detection of nonTor Traffic. Journal of Cyber Security, Vol. 6.2, pp 171–194, November 2017. doi: 10.13052/jcsm2245-1439.624