Artificial Intelligence methods suitable for Incident Handling Automation

Roumen Trifonov[1*], *Radoslav* Yoshinov[2], *Slavcho* Manolov[1], *Georgi* Tsochev[1]and *Galya* Pavlova[1]

[1]Faculty of Computer Systems and Technologies, Technical University of Sofia, 1000 Sofia, Bulgaria,
[2]Laboratory of Telematics, Bulgarian Academy of Sciences, 1000 Sofia, Bulgaria

Abstract. The Faculty of Computer Systems and Technology at Technical University – Sofia undertook analyses and experiments on the use of Artificial Intelligence methods in the field of Information Security. In the course of the study a Cyber Defence related to the classification consisting of three phases has been proposed: Operative Cyber Intelligence, Tactical Cyber Intelligence and Incident Handling. It has been found that there is no universal Artificial Intelligence method effective for all phases mentioned above and for all applications. In each case, a set of criteria should be developed to select (and then experiment) an appropriate method (or combination of methods). A selection of methods for the first two stages of Cyber Defence was described in previous work. In the present paper are considered some of the considerations related to effective use of appropriate methods for the Incident Handling phase.

1 Introduction

The Faculty of Computer Systems and Technologies at the Technical University-Sofia conducts for several years analyses and experiments on the implementation of Artificial Intelligence methods in the field of Information Security [22]. Currently, these studies are funded by Bulgarian National Science Fund in the frameworks of the project "Increasing the level of the Network and Information Security using Artificial Intelligence methods".

In the course of the study, we came to two fundamental conclusions, which (at least on the referenced so far sources) have not been formulated in an explicit form:

A. The Cyber Defence (depending on objectives and applied methods and tools) can be divided into three components:
- immediate coverage of attacks - we refer it to the so-called Tactical Cyber Intelligence;
- anticipating the actions of the possible adversary - refers to the so-called Operational Cyber Intelligence and
- removal of the consequences of the attack - refers to the so-called Incident Handling.

B. We have found that there is no universal Artificial Intelligence method that is effective for all phases mentioned above and for all applications. In each case, a set of criteria should be developed to select (and then experiment) an appropriate method (or combination of methods). Moreover, the type of detection depends on the nature of the threats (knowns, unknowns and combinations of the two types) [1].

The present article reflects the next step in the development of the above-mentioned project, following the philosophy of assessing the capabilities of existing Artificial Intelligence methods to resolve some or other cyber-security issues with a high probability of efficiency. Comparative analysis above all should focus on the ability of these methods to minimize false positive or negative results, given that often the result of countering a false-identified attack can have harmful consequences comparable to a real attack.

The previous articles [2,3,4,5,6,7,8,9,10,11,12] reporting on the works implemented under the project are devoted to the first two (on the above-mentioned classification) stage of the Cyber Defence: the Tactical Cyber Intelligence, where we have used Multi-Agent system of self-learning agents and Operational Cyber Intelligence with Echo State Neural Networks plus Reservoir Computing.

2 Cyber Security Incident Handling

Incident response has become necessary because attacks frequently cause the compromise of personal and business data. Incidents involving viruses, worms, Trojan horses, spyware, and other forms of malicious code have disrupted or damaged millions of systems and networks around the world. Heightened concerns about national security and exposure of personally identifiable information (PII) are also raising awareness of the possible effects of computer-based attacks. These events—and many more—make the case daily for responding quickly and efficiently when computer security defences are breached. Besides the business

[*] Corresponding author: r_trifonov@tu-sofia.bg

reasons to establish an incident response capability, the organizations must comply with law, regulations, and policy directing a coordinated, effective defence against information security threats. It is important that organizations have a formal, focused, and coordinated approach to responding to incidents. To effectively implement such a capability, an organization should have an incident response plan. The plan provides the organization with a roadmap for implementing its incident response capability. The plan should provide a high-level approach for how the incident response capability fits into the overall organization. Each organization needs a plan that meets its unique requirements, which relate to the organization's mission, size, structure, and functions. The plan should lay out the resources and management support that is needed to effectively maintain and mature an incident response capability.

At the current level of threats and attacks the main field of battle is transferred to specialized units with highly qualified specialists –so called Computer Security Response Teams (CERT-s). That's why the new complex relationship arise that require precise regulation. The Incident Handling as a formal procedure is governed by several international standards including: Recommendations E.409 and H.1500 of ITU (International Telecommunication Union), ISO 18044 and SR 800-61 of NIST (National Institute for Standardizations and Technologies).

The European Network and Information Security Agency (ENISA) defined cyber security incident as "an IT disruption that limits or eliminates the expected availability of services, and/or is the unauthorized publication, acquisition and/or modification of information", whereby "single or a series of unwanted or unexpected information security events make a significant probability of compromising business operations and threatening information security. A cyber security incident can involve a real or suspected breach or the unlawful act of exploiting vulnerability." [13].

According to [14] "The concept of computer security incident handling has become widely accepted and implemented. One of the benefits of having an incident handling capability is that it supports responding to incidents systematically (i.e., following a consistent incident handling methodology) so that the appropriate actions are taken. Incident handling helps personnel to minimize loss or theft of information and disruption of services caused by incidents. Another benefit of incident handling is the ability to use information gained during incident handling to better prepare for handling future incidents and to provide stronger protection for systems and data."

The incident response process has several phases [14], [15], [16]. The initial phase involves establishing and training an incident response team, and acquiring the necessary tools and resources. During preparation, the organization also attempts to limit the number of incidents that will occur by selecting and implementing a set of controls based on the results of risk assessments. However, residual risk will inevitably persist after controls are implemented. Detection of security breaches is thus necessary to alert the organization whenever incidents occur. In keeping with the severity of the incident, the organization can mitigate the impact of the incident by containing it and ultimately recovering from it. During this phase, activity often cycles back to detection and analysis—for example, to see if additional hosts are infected by malware while eradicating a malware incident. After the incident is adequately handled, the organization issues a report that details the cause and cost of the incident and the steps the organization should take to prevent future incidents. The figure 1 illustrates the incident response life cycle.

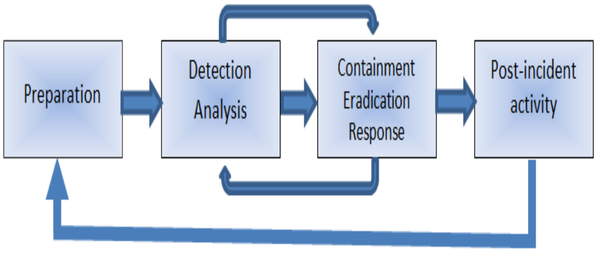

Fig. 1. Incident response life cycle

The essential features of this process are indicated at [16]: "Efficient handling of all the details of an incident is essential to solving security problems. The ability to track the status of an incident and to combine collected input, research results and information about the actions taken helps lead to incident resolution. Teams should use a tool for organizing and tracking user reports and questions, incidents and actions taken together with specific issue (incident) status e.g. opened, delayed, in process, solved."

3 Incident Handling Automation

Obviously, the minimization of the reaction time for Incident Handling can minimize its consequences and can be one of the main factors for effectiveness of this process. A lot of steps have been taken in this direction in recent years:

- in the direction of the Automation of information exchange between involved parties (Fig. 2) by structured machine-processed messages specialized languages for data exchange have been introduced, such as IODEF (Incident Object Description Exchange Format) - specification of the IETF Working Group;
- in the direction of Automation of elements of the process workflow various techniques have been widely applied, such as so-called Trouble Ticket (or accident report), which is a mechanism for describing the incident in a unified way so as to ensure its identification, reporting, processing and resolution. It is specified by the IETF (Internet Engineering Task Force) Recommendation RFC 1297 [17] and is similar to a "patient card" in a hospital.

Incident detection and analysis would be easy if every precursor or indicator was guaranteed to be accurate; unfortunately, this is not the case. For example, Intrusion Detection systems may produce false positives—incorrect indicators.

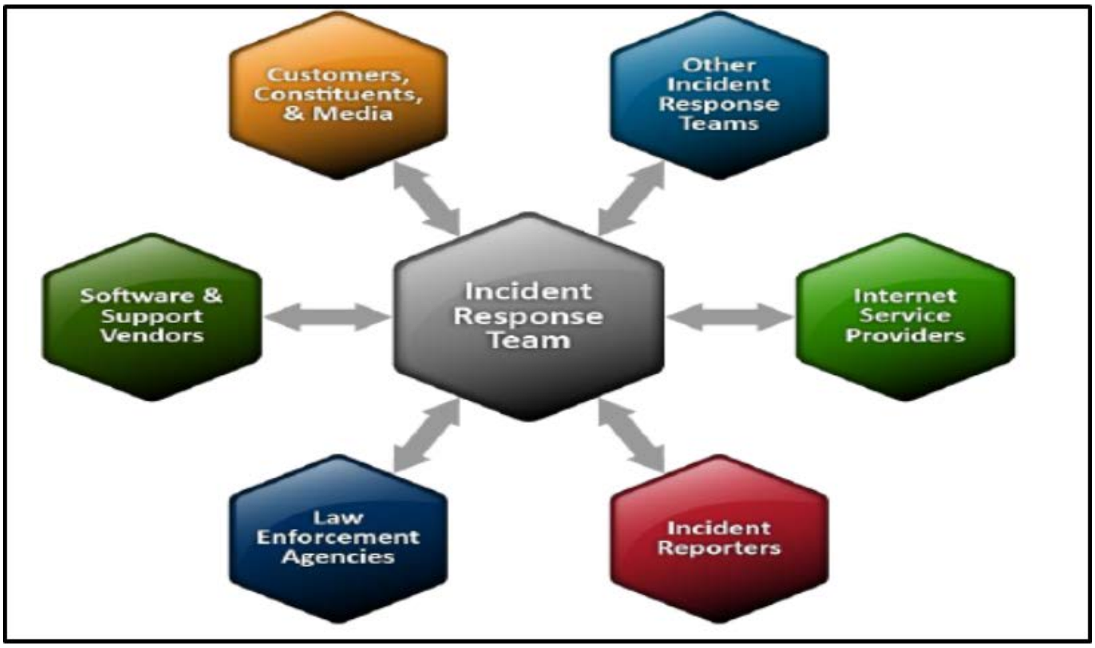
Customers, Constituents, & Media
Other Incident Response Teams
Software & Support Vendors
Incident Response Team
Internet Service Providers
Law Enforcement Agencies
Incident Reporters

Fig. 2. Direction of the Automation of information exchange between involved parties

This demonstrates what makes incident detection and analysis so difficult: each indicator ideally should be evaluated to determine if it is legitimate. Making matters worse, the total number of indicators may be thousands or millions a day. Finding the real security incidents that occurred out of all the indicators can be a daunting task.

Even if an indicator is accurate, it does not necessarily mean that an incident has occurred. Some indicators, such as a server crash or modification of critical files, could happen for several reasons other than a security incident, including human error. Given the occurrence of indicators, however, it is reasonable to suspect that an incident might be occurring and to act accordingly. Determining whether a particular event is actually an incident is sometimes a matter of judgment. In many instances, a situation should be handled the same way regardless of whether it is security related.

Some incidents are easy to detect, but the majority of incidents are not associated with such clear symptoms. In incident handling, detection may be the most difficult task. Incident handlers are responsible for analyzing ambiguous, contradictory, and incomplete symptoms to determine what has happened. Although technical solutions exist that can make detection easier, the best remedy is to build a team of highly experienced and proficient staff members who can analyze the precursors and indicators effectively and efficiently and take appropriate actions, following a pre-defined process and documenting each step taken.

As has been said, the consequences of the incident directly depend on the speed of the process of Incident Handling. Therefore, at present, the main goal of the different forms of automation is to minimize the time to deal with the incident. That’s why in our analysis we have started the researches with the most time-sensitive (in our opinion) element - the Triage, which consists of three sub-phases: Detection, Initial Classification and Assignment.

4 Artificial Intelligence methods

Not so much of the scarce literary sources describing attempts to apply Artificial Intelligence methods in Incident Handling [18, 19], but on the base of our experience of introduction of Artificial Intelligence methods in Tactical, and above all, Operational Cyber Intelligence [11, 12], we have come to the conclusion that at present the main function of Artificial Intelligence in Incident Handling can be solving a classification task, i.e. the unambiguous reference of current incident to one of the elements of the Classification Scheme, where for each element relevant procedures and workflows have been developed.

In addition, experiments on the application of Artificial Intelligence methods in Operational Cyber Intelligence have shown that the most important part of solving this classification task is to find so-called "features", i.e. characteristics that adequately reflect objective dependencies on the classification status. The “feature” extraction can be defined as an operation which transforms one or several characteristics into a “feature vector”. Identifying and extracting good “features” from all characteristics is a crucial step, because otherwise the classification algorithm will have trouble identifying the class of these “features”.

If we stick to this approach, it should be noted that the application of Artificial Intelligence to the Incident

Handling must be solved both by right and by the opposite task.

The right task assignment can be briefly summarized as follows: to be found the features in the attributes contained in the unified incident reports based on international standards (for example, Trouble Ticket).

The reverse task appears to be more difficult - it consists in changing the attributes of unified incident reports according to international standards so that they can reflect more adequately the affiliation of a particular incident to one of the elements of the classification scheme. Here we can say that this task is not yet considered relevant by the project team.

The analysis of relatively scarce literary sources and the experience of the implementation of Artificial Intelligence methods in Tactical Cyber Intelligence directed us to so called Reinforcement Learning method. The essence of Reinforcement Learning is training through interaction. A Reinforcement Learning agent interacts with its environment and, upon observing the consequences response to rewards received. This paradigm of trial-and error learning has its roots in behavior psychology, and is one of the main foundations of Reinforcement Learning. The other key influence on this method is optimal control, which has lent the mathematical formalisms (most notably dynamic programming) that underpin the field.

The best sequence of actions is determined by the rewards provided by the environment. Every time the environment transitions to a new state, it also provides a scalar reward Rt+1 to the agent as feedback. The goal of the agent is to learn a policy (control strategy) that maximizes the expected return (cumulative, discounted reward) (Fig. 3).

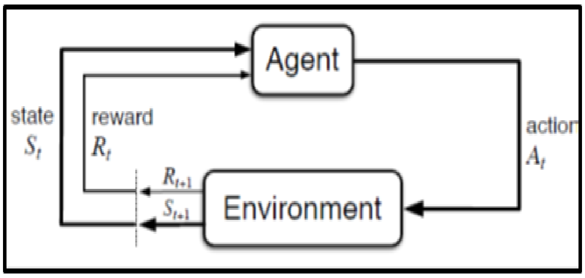

Fig.3. Environment transitions to a new state

Unlike the Controlled Learning usually implemented in Neural Networks, Reinforcement Learning is realized using previously collected examples or a set of data for training that is not suitable for Interactive Learning. That’s why the bulk of the training can be accomplished by analyzing a collection of existing incidents, identifying key attributes that have patterns of correlation to categories, and creating a model to make predictions from these patterns. In this situation, the main purpose of the agent is to maximize the remuneration achieved in the long run, i.e. the sum of the awards received from all situations or conditions that will be reached in the future:

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots . = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \quad (1)$$

where r is a consequence of an action that results in a digital reward for each time step and Y represents the reported discount rate to show how important the future reward is.

For the cases where the model is unable to predict a value with high enough confidence, the result with this incident can be later manually corrected by a human, the change event must be collected and used to improve the model (Fig. 4).

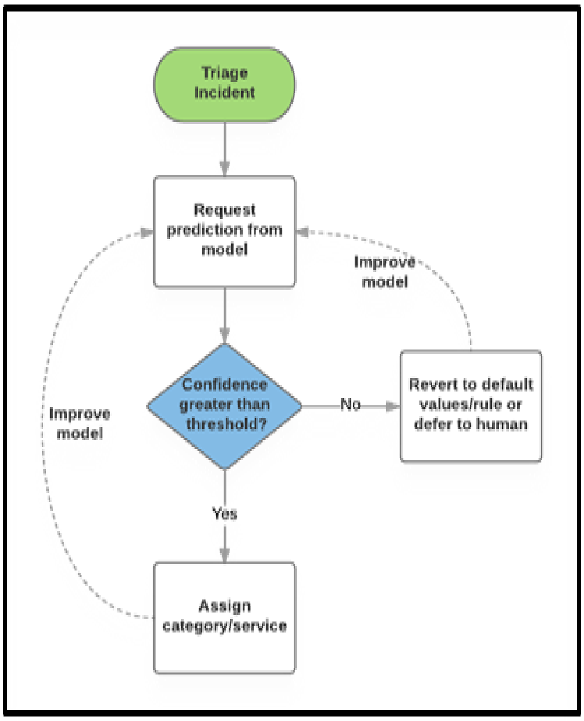
Triage Incident
Request prediction from model
Improve model
Confidence greater than threshold?
No
Revert to default values/rule or defer to human
Improve model
Yes
Assign category/service

Fig. 4. Algorithm for improvement of the prediction model

5 Conclusion

The state of the art of the works, described in this article, can be defined as a development of a theoretical model. Experiments are yet to come and that's why it is necessary to point out it is still too early to declare any definitive conclusions. It is quite possible that the hypothesis is not quite correct and the process of Reinforcement Learning can be not convergent.

As can be seen from the above, in the global practice processes of introducing Artificial Intelligence methods at the different levels of Cyber Defense are at very different stages: while in the Cyber Intelligence they have long gone out of the phase of research and experiments and are used for building real effective systems, in the field of Incident Handling these studies are in a very initial phase and require commitment of substantial resources. Furthermore, the question arises as to the probable application of possible outcomes of Cyber Iintelligence in the activity of Incident Handling systems, which are intended to speed up the identification of the incident classification.

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. National Cyber Security Strategy “Cyber Resilient Bulgaria” Sofia, 2016
2. R. Trifonov, S. Manolov Application of multi-agent systems for network and information protection Proceedings of the International Conference on Information Technologies (InfoTech-2014) 18 - 19 September 2014, Varna, Bulgaria
3. R. Trifonov, G. Tsochev, S. Manolov, Radoslav Yoshinov, G. Pavlova, A Survey of Artificial Intelligence for Enhancing the Information Security, Int. J. of Development Research, **07**, 11, November (2017), pp.16866-16872, ISSN: 2230-9926
4. R. Trifonov, S. Manolov, R. Yoshinov, G. Tsochev, G. Pavlova, Artificial Intelligence Methods for Cyber Threats Intelligence, Int. J. of Computers, **2** (2017) pp. 129-135, ISSN: 2367-8895,
5. R. Trifonov, G. Tsochev, R. Yoshinov, S. Manolov and G. Pavlova. Conceptual model for cyber intelligence network security system, Int. J. of Computers, **11** (2017) pp. 85-92,ISSN: 1998-4308
6. R. Trifonov, S.Manolov , R. Yoshinov , G. Tsochev, G. Pavlova. An adequate response to new Cyber Security challenges through Artificial Intelligence methods. Applications in Business and Economics, WSEAS Transactions on Business and Economics, **14** (2017) pp. 272 - 281, E-ISSN: 2224-2899
7. R. Trifonov, G. Tsochev, S. Manolov, R. Yoshinov, G. Pavlova, Increasing the level of network and information security using artificial intelligence, *Fifth Intl. Conf. Advances in Computing, Communication and Information Technology- CCIT* 2-3 September (2017) Zurich, Swiss, ISBN: 978-1-63248-131-3
8. R. Trifonov, G. Tsochev, G. Pavlova, R. Yoshinov, S. Manolov, Adaptive Optimization Techniques for Intelligent Network Security, *4th International Conference on Mathematics and Computers in Sciences and Industry MCSI* (2017), August 24-26, Corfu Island, Greece, Conference Publishing Services of IEEE
9. R. Trifonov, G. Tsochev, R. Yoshinov, S. Manolov, G. Pavlova, Conceptual model for cyber intelligence network security system, Int. J. of Computers, **11 (**2017) ISSN: 1998-4308
10. R. Trifonov, O. Nakov, P. Vatchkov, S. Manolov, R. Yoshinov, G. Popov, G. Tsochev, G. Pavlova. Intelligent methods and Cybersecurity, *XXV Conference Telecom* (2017) 26-27 October, NSTC, Sofia, Bulgaria, p. 113-120
11. R. Trifonov, S. Manolov, R. Yoshinov, G. Tsochev, S. Nedev, G. Pavlova, Operational Cyber Threat Intelligence supported by Artificial Intelligence methods. *Proceedings of the International Conference on Information Technologies (InfoTech-2018)* 20 - 21 September (2018) Varna, Bulgaria
12. R. Trifonov, S. Manolov, R. Yoshinov, G. Tsochev, G. Popov, G. Pavlova, New Approaches in the Examination of the Cyber Threats. *Proceedings of the International Conference on Information Technologies (InfoTech-2018)* 20 - 21 September (2018) Varna, Bulgaria
13. *ENISA Threats Landscape Report 2016*: 15 Top Cyber-Threats and Trends, ENISA, January (2017)
14. *Computer Security Incident Handling Guide Special Publication 800-61* Revision 2 NIST, August (2012)
15. *Good Practice Guide for Incident Management*, ENISA (2010)
16. *Strategies for Incident Response and Cyber Crisis Cooperation* Version 1.1, ENISA, August (2016)
17. *Request for Comments 1297* “Internal Integrated Trouble Ticket System – Functional Specification Wishlist” Merit Network, Inc. January (1992)
18. R. Young *AI-driven automation for Incident Management*, Astound, March 8, (2017)
19. *Security Team’s Operational Requirements Milestone* MS2.4.2. GN3-10-073 GEANT (2014)
20. R.S. Sutton *Reinforcement Learning An Introduction* Cambridge University Press (1998)
21. K. Arulkumaran, M.P. Deisenroth, M. Brundage, A.A. Bharath A Brief Survey of Deep Reinforcement Learning IEEE Signal Processing Magazine Special Issue on Deep Learning for Image Understanding Nov. (2017)
22. G. Popov and K. Raynova, Diversity in nature and technology — Tool for increase the reliability of systems, *15th International Conference on Electrical Machines, Drives and Power Systems (ELMA)*, Sofia (2017)