ATLANTIS PRESS Journal of Robotics, Networking and Artificial Life, Vol. 4, No. 4 (March 2018) 303-307

Artificial Immune Ecosystems: the role of expert-based learning in artificial cognition

Pierre Parrend

ICube laboratory, University of Strasbourg, UMR CNRS 7357, Strasbourg, France, ECAM Strasbourg-Europe, Schiltigheim France, pierre.parrend@unistra.fr

Fabio Guigou, Julio Navarro, Aline Deruyver, Pierre Collet

ICube laboratory, University of Strasbourg, UMR CNRS 7357, Strasbourg, France

Abstract

The rapid evolution of IT ecosystems significantly challenges the security models our infrastructures rely on. Beyond the old dichotomy between open and closed systems, it is now necessary to handle securely the interaction between heterogeneous devices building dynamic ecosystems. To this regard, bio-inspired approaches provide a rich set of conceptual tools, but they have failed to lay the basis for robust and efficient solutions. Our research effort intends to revisit the contribution of artificial immune system research to bring immune properties: security, resilience, distribution, memory, into IT infrastructures. Artificial immune ecosystems support a comprehensive model for anomaly detection and characterization, but their cognitive capacity are limited by the state of the art in machine learning and the rapid evolution of cybersecurity threats so far. We therefore propose to enrich the cognitive process with expert-based learning for reinforcement, classification and investigation. Application to system supervision using system logs and supervision time series confirms the relevance and performance of this model.

Keywords: Artificial Immune Systems, Cybersecurity, Immunity, Computational Ecosystem, Anomaly detection

1. Introduction

Old closed IT infrastructures increasingly turn into open interconnected ecosystems of heterogeneous devices: this shift is put to light in particular by the numerous challenges posed by IoT Security [1]. In this context, the detection of anomalies caused by cyberattacks becomes an ever more tedious task to identify known anomalies, new 'zero-day' anomalies or signs of system instability following these anomalies. Bio-inspired architectures and algorithms provide to this regard a promising set of concepts and solutions for multi-scale dynamic ecosystems. However, no robust and efficient solution exists to solve the issue of generic anomaly detection so far. We therefore work on a new generation Artificial Immune Systems (AIS) called Artificial Immune Ecosystem (AIE): it encompasses not only the analysis layer as classical AIS do, but also a basis layer for robust and resilient data management, an immune communication protocol [2] as well as a supervision and knowledge management layer [3]. The AIE shall embed the properties of natural immune systems: security, resilience, distribution, memory. To this aim, their cognition capacity for anomaly identification and characterization in a context of dynamic systems and rapidly evolving threats is a key requirement. We are convinced that cognition, *i.e.* the support of search, adaptability, memory and learning [4] in such heterogeneous environments will not be achieved in a short time span through fully automated solution, and that the contribution of human creativity in the analysis process is a major enabler for adaptation and learning. We therefore propose in this paper to extend the AIE with expert-based learning for reinforcement, classification and investigation. We apply this model to system supervision using system logs and supervision time series to confirm its relevance and performance.

Section 2 introduces related works. Section 3 defines the role of cognition in the Artificial Immune Ecosystem (AIE). Section 4 formalizes the expert-aided learning in the AIE. Section 5 presents the experiments and the evaluation of this model. Section 6 concludes this work.

2. Related Work

2.1. *Mechanisms of natural immunity*

It is possible to summarize natural immunity as the perception of three types of signals by the immune system [5]:
1-signal: the affinity between antigens and antibodies enable these latter to identify known pathogens. This process typically occurs through B-cells and T-cells in acquired immunity.
2-signals: the affinity between immune receptors and PAMPs (Pathogen-Associated Molecular Patterns, such as endotoxin from certain bacteria) triggers infectious inflammation and multiplies the immune reaction. This process typically occurs through Pattern Recognition Receptors (PRR) in innate immunity.
3-signals: the affinity between immune receptors and DAMPs (Danger Associated Molecular Patterns, such as signs of necrosis or intracellular proteins like DNA when released in the intercellular milieu) triggers non-infectious inflammation. This process typically occurs through TLR (Toll-Like Receptors, a specific kind of PRR) in innate immunity.

2.2. *Cognition and immunity*

Natural immunity is a reference model of learning and cognition where a fully decentralized and autonomous system supports a comprehensive loop of information input, processing and subsequent reaction [6]. This loop is known since Varella as the enactment principle [7]. The cognitive domain covers the range of search, adaptability, memory and learning [4] capabilities. Search is the capability of processing input information for partially hidden signals. Adaptability is the capability of letting information processing evolve over time to cope with environment evolution. Memory enables to retain raw or processed information about passed events. Learning occurs when memory content trigger the adaptation of search capability. It can concern both facts and behaviours.
Bourgine and Steward propose a more global definition of cognition: “A system is cognitive if and only if sensory inputs serve to trigger actions in a specific way, so as to satisfy a viability constraint” [8], *i.e.* if it is capable of reaction in a context where its viability is at risk.

Although these properties have been devised in a theoretical context, they prove to be key to the construction of an efficient Artificial Immune Ecosystem for multi-scale IT infrastructures.

3. Cognition and the Artificial Immune Ecosystem

3.1. *The artificial immune process*

A major difference exists between natural immunity and artificial immune processes: In natural systems, the perception of the three (3) signal types occurs through the various immune cell types in parallel. In artificial systems, detection mechanisms are best defined as a process involving several analysis approaches so far, as defined in the artificial immune detection process in Figure 1. 1) The first step for anomaly detection is deterministic memory-based recognition of pathogens. It typically uses explicit rules with a tightly define scope, such as in antiviruses. 2) The second step consists in stochastic immune detection: 2.1) detection based on known antigens, for instance using fuzzy detectors, similar to the 1-signal model of natural immunity, 2.2) recognition of a generic model similar to the 2-signals model and 2.3) detection of danger signal similar to the 3-signals models. 3) The third step consists in the characterization of the anomaly. It can be automated or assisted by the expert as specified in Section 4. 4) Next, the system memory is updated to leverage the knowledge gained through the characterization step. 5) A reaction is performed – which is beyond the scope of this contribution. One should note the difference between step 1) which is deterministic and well understood and step 2.1) which leverages stochastic technologies and is an active research field.

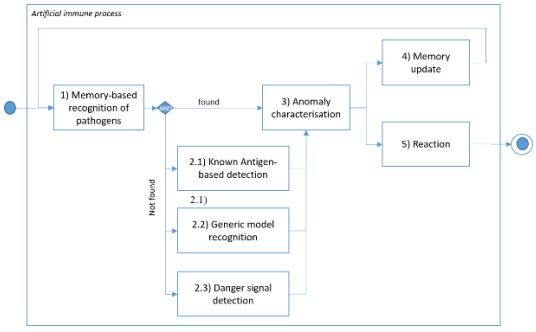

Figure 1 : The artificial immune detection process

On can note several radical differences between the natural immune system (IS) and the artificial immune system process:
1) Artificial immunity greatly benefits of integrating deterministic detectors since such technologies are already available and mature, whereas natural systems rely on probabilistic affinities only.
2) In natural IS, no notion of process flow exists, immune mechanisms occur in parallel, whereas an artificial process embeds expert knowledge in its own design.

3.2. *Anomaly characterization*

The step 2.2) of anomaly characterization entails an automated step and a manual-driven step, as shown in Figure 2. We share the vision of D. Engelbart that in highly evolutive environments, interactive systems which "increas[e] the capability of a man to approach a complex problem situation, to gain comprehension to suit his particular needs, and to derive solutions to problems" [9] are a necessity to overcome the intrinsic limitations of bottom-up fully automated systems. Step 3.1) consists in automated detection, typically using statistical or machine learning approaches. If the analysis process provides a satisfactory output here, anomaly characterization is completed. Otherwise, the output of this automated step is presented to the expert for further analysis. If analysis requires a manual confirmation of automated classification for reinforcement, the expert validates or invalidates the system proposal (step 3.2.1). If the analysis requires a manual classification, the expert allocates the anomaly to a category (1 dimension), potentially creating a new one (step 3.2.2). If the characterization is more complex (2 or more dimensions), the expert performs a deeper investigation (step 3.2.3).

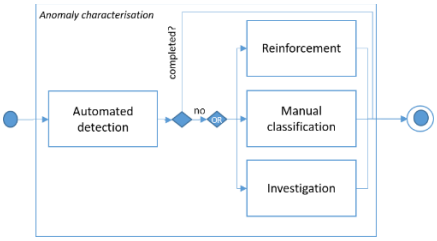

Figure 2 : Anomaly characterisation

The output of steps 3.1, 3.2.1 and 3.2.2 enables the continuous adaptation of the learning artificial immune process. Step 3.2.3 requires that the process can embed complex knowledge in latter analyses such as new detectors, new algorithms or new algorithm parameters.

4. The role of the Expert in the Artificial Immune Ecosystem

The expert intervenes in anomaly characterization for reinforcement, classification of investigation.

4.1. *Expert-based reinforcement (step 3.2.1)*

In expert-based reinforcement, the expert approves or invalidates the classification into a given category.

$$prop = \{S_U \xrightarrow{I} C_i\} \tag{1}$$

$$prop \overset{\mathrm{E}}{=} true \vdash S_U \xrightarrow{E} C_i \vdash S_U \in C_i \tag{2}$$

$$prop \overset{\mathrm{E}}{=} false \vdash S_U \notin C_i \tag{3}$$

As stated in Equation 1, a classification proposal *prop* that a signal S_U pertains to a given category C_i is made by the system. If the expert labels this classification as correct, the membership of S_U to C_i is confirmed, as given in Equation 2. If the expert labels this classification as incorrect, the membership of S_U to C_i is invalidated, as given in Equation 3.

4.2. *Expert-aided classification (step 3.2.2)*

In expert-aided classification, the expert provides the category of unknown items, when the system does not manage to find a satisfactory category itself.

$$K = \{C_i\} \tag{5}$$

$$C_i \in \{A_0, \ldots, A_n, AAP_0, \ldots, AAP_m, D_0, \ldots, D_o\} \tag{6}$$

$$S_U \xrightarrow{E} C_i \vdash S_U \in C_i \tag{7}$$

$$S_U \xrightarrow{E} C_u \in \{A_u, AAP_u, D_u\} \vdash S_U \in C_u \wedge K = \{K \cup C_u\} \tag{8}$$

Equation 5 states that system knowledge K is compound of a set of known anomaly categories $\{ C_i \}$. Equation 6 states that this set of known anomaly categories entails a set of n traces of Anomalies $\{A_i\}$ [2], a set of m traces of Anomaly Accompanying Pattern $\{AAP_i\}$ [3], a set of o traces of danger $\{D_i\}$ [2]. Equation 7 states that when this signal S_U is affected by the expert ($\xrightarrow{E}$) to an existing anomaly category C_i which can be either a known anomalous event A_i, *a* known pattern AAP_i or a known danger signal D_i symptom of abnormal system use, this existing category C_i is extended with anomaly U. This process occurs when C_i does not contain U yet. Equation 8 states that when this signal S_U is affected by the expert to a new category C_n, which is either an anomalous event A, a pattern AAP or a danger signal D, the set C of

anomaly categories is extended with C_n, that entails a single anomaly *U* at this point.

4.3. *Expert investigation (step 3.2.3)*

In expert investigation, the experts performs the characterization of the properties of an unknown item. This characterization can typically lead to the definition of new detectors.

$$S_U \xrightarrow{EI} S_k = \{p_i\} \quad (9)$$

$$S_k \xrightarrow{ED} \mathcal{D}_K \quad (10)$$

$$K = \{K \cup S_k\} \wedge \mathcal{D} = \{\mathcal{D} \cup \mathcal{D}_K\} \quad (11)$$

As stated in Equation 9, expert investigation *EI* leads to the analysis of an unknown signal S_U as a set of anomaly properties $\{p_i\}$, which then becomes a known signal S_k. As stated in Equation 10, the intervention of expert design process *ED* can lead to the creation of a detector D_K for the signal S_k. As stated in Equation 11, the knowledge base *K* is then extended with signal S_k, and the detector base *D* with detector D_K. Note that in the case of investigation, the knowledge is compound of the subproperties $S_k = \{p_i\}$ of each anomaly, rather than out of anomaly categories as in the previous cases of classification. Expert-aided classification is thus a special case of the investigation where $|S_k = 1|$ and $S_k \in \{Ci\}$, *i.e.* the properties can be assigned to an explicit, well-known category.

5. Evaluation

5.1. *A cognitive Artificial Immune Ecosystem*

In the frame of Bersini's definition of cognition domains [4], the proposed artificial immune process is actually a cognitive process: it supports search, adaptability, memory and learning, and does so for ensuring the viability of the system. Steps 1) deterministic memory-based recognition of pathogens and 2) stochastic recognition of pathogens embed the search facility. Step 3) Anomaly characterization embeds adaptability. Step 4) Memory update ensures the memory property together with its exploitation in steps 1), 2), 3). The combination of search, adaptability and memory ensures the capability of learning facts; step 3.2.3) investigation ensures the capability of learning behaviours. In our current model, the human expert supports the reaction process, the viability constraints of IT systems equipped with an AIE and does most of the adaptability capability and behavior learning. The expert-augmented AIE is thus a cognitive system in the sense of Bourgine's definition [8], but immense research questions are open before the community can transform the proposed process into a fully autonomous one.

5.2. *Expert-based reinforcement*

We studied the efficiency of expert-based reinforcement with the Morwilog tool [10], which exploits expert feedback to validate the identification of sequences of system logs as occurrences of multi-step attacks. Although these results are theoretical, they provide highly promising insights *wrt.* the capability of handling rapidly evolving complex signals in the context of cybersecurity.

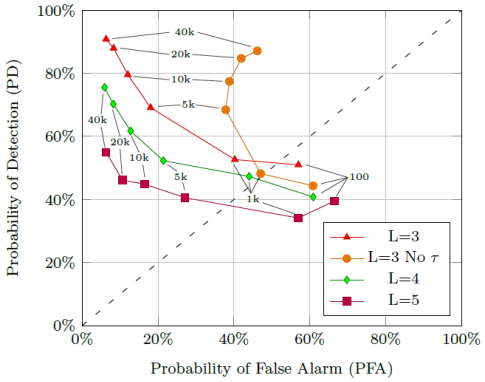

Figure 3 : Evaluation of expert-based reinforcement in Morwilog

5.3. *Expert-aided classification*

We studied the efficiency of expert-aided classification in the application of the Artificial Immune Ecosystem to network supervision [10], in the particular case of the analysis of time series information. Again, the theoretical results show that the involvement of an expert in the learning process, here for classification, greatly quickens the analysis capability. On a dataset of 10.000 entries, 30 requests on expert feedback, *i.e.* 0,3%, enables to achieve 89,7% of precision as given by classification F-score. 500 expert feedbacks, *i.e.* 5%, provides 99% precision.

Maximum expert feedback	20	30	50	100	500	Unlimited
F-score	.36	.70	.72	.72	.78	.78

Table 1: Impact of expert-based classification

5.4. *Expert investigation*

We studied the efficiency of expert investigation in the context of Morwilog tool: before the reinforcement phase can start, the AIE provides suspicious traces as input to the expert to enable him to characterize prototypical multi-step attacks. Based on a list of suspicious traces, the expert thus identifies the actual threat and confirms that a given event sequence builds a single attack.

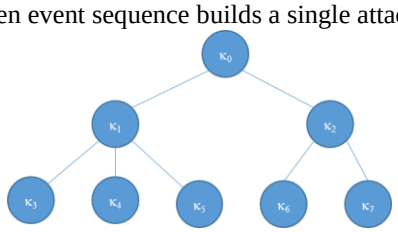

Figure 4 : Attack steps identified through Morwilog investigation

6. Conclusions and perspectives

In this contribution, we propose a process for enforcing artificial immunity, *i.e.* a bio-inspired model supporting the search for three immune signals: pathogens, pathogen-associated patterns and danger. The artificial immune process entails rich detection capabilities, as well as anomaly characterization, which is key to ensure adaptability and memory, and thus learning. In the context of a rapidly evolving environment, we show that involving the expert in this process for reinforcement, classification or investigation provides significant improvements in the learning process, which can then be considered a cognitive one.

Acknowledgements

We express our thanks to the colleagues who have contributed to projects in which we developed the Artificial Immune Ecosystem: Véronique Thomas-Vaslin, Véronique Legrand. This work was founded by ICube Laboratory (SENSAI project), French Banque Publique d'Investissement (BPI) under AAP-19 HuMa project, Agence Nationale de la Recherche Technologique (ANRT) and the enterprise IPLine.

References

1. S. M. Hashemi and J. He. La-based approach for iot security. *Journal of Robotics Networking and Artificial Life*, 3(4):240–248, 2017.
2. P. Parrend, P. David, F. Guigou, C. Pupka, and P. Collet. The AWA artificial emergent awareness architecture model for artificial immune ecosystems. *In IEEE Congress on Evolutionary Computation 2017*, Special Session on Artificial Immune Systems: Algorithms, Simulation, Modelling & Theory, June 2017.
3. F. Guigou, P. Parrend, and P. Collet. An artificial immune ecosystem model for hybrid cloud supervision. *In Complex System Digital Campus'15*, Sep 2015.
4. H. Bersini and F. J. Varela. Hints for adaptive problem solving gleaned from immune networks. *In International Conference on Parallel Problem Solving from Nature*, pages 343–354. Springer, 1990.
5. J. Greensmith, U. Aickelin, and S. Cayzer. Detecting danger: The dendritic cell algorithm. *In Robust Intelligent Systems*, pages 89–112. Springer, 2008.
6. J. E. Hunt and D. E. Cooke. Learning using an artificial immune system. *Journal of network and computer applications*, 19(2):189–212, 1996.
7. F. Varela. Invitation aux sciences cognitives, 1988.
8. P. Bourgine and J. Stewart. Autopoiesis and cognition. *Artificial life*, 10(3):327–345, 2004.
9. D. C. Engelbart. Augmenting human intellect: A conceptual framework. Summary report afosr-3223 under contract af 49 (638)-1024, SRI project 3578 for air force office of scientific research. Stanford Research Institute. Retrieved March, 1:2007, 1962.
10. J. Navarro Lara, A. Deruyver, and P. Parrend. Morwilog: an ACO-based system for outlining multi-step attacks. *In IEEE Symposium Series on Computational Intelligence* (IEEE SSCI 2016). IEE, Dec 2016.
11. F. Guigou, P. Collet, and P. Parrend. The artificial immune ecosystem: a bioinspired meta-algorithm for boosting time series anomaly detection with expert input. In EvoApplications, *20th European Conference on the Applications of Evolutionary Computation*, Apr 2017.