ARTIFICIAL IMMUNITY-BASED CORRELATION SYSTEM

Guillermo Suarez-Tangil, Esther Palomar, Sergio Pastrana, Arturo Ribagorda
Department of Computer Science, University Carlos III of Madrid, Avda. Universidad 30, 28911 Madrid, Spain
gtangil@pa.uc3m.es, epalomar@inf.uc3m.es, spastran@inf.uc3m.es, arturo@inf.uc3m.es

Keywords: Artificial Immune System; Event Correlation; Security Event Information Management System; Intelligent Rule Generation; Adaptive System.

Abstract: Security information event management (SIEM) technologies focus on developing effective methods and tools to assist network administrators during the whole network security management. Though there is a vast number of novel initiatives and contributions in providing adaptiveness and intelligence in this research field, there are still many problems that need be solved. In particular, event correlation are currently emerging as an essential field to be optimized specially due to the widespread adoption of botnets to launch attacks. This position paper explores the biological immune system's characteristics of learning and memory to solve the automatic generation of event correlation rules by applying Artificial Immune Systems (AISs).

1 INTRODUCTION

Nowadays, network security management has to deal with two main critical tasks: On the one hand, different security data sources (known as sensors), e.g. intrusion detection systems (IDSs), firewalls, server logs, to name a few produce high amounts of heterogeneous information, generally difficult to understand. Moreover attackers may use IDS-stimulators to disguise their intrusions by hiding attacks into an alert storm (Mutz et al., 2003). On the other hand, modern attacks pose a challenge to the network infrastructure as these attacks may be not noticed when they are inspected separately.

To encounter these challenges, cooperation among these sensors becomes essential for the former, whilst event correlation appears as the best palliative for the latter. Security Information and Event Management (SIEM) systems, as a holistic solution, help to gather, organize and correlate security network information as well as reducing the amount of time spent by security administrators. In particular, OSSIM (OSSIM, 2010) is an open source SIEM implementation which centralizes the detection and monitoring of the security events within an organization.

1.1 Problem Statement

A typical SIEM architecture facilitates experts to supervise the security status of an organization. However, current SIEM systems lack of an efficient mechanism to generate correlation rules and cannot adaptively predict novel attacks either (Anuar et al., 2010). Furthermore, the continuous evolution of attacks, specially recent distributed, multi–step attacks pose an additional challenge to experts and also the entire detection process.

By providing an intelligent adaptability to the correlation engine, we envision that time spent on detecting zero-day attacks can be significantly reduced. For instance, various works focused on the application of Artificial Intelligence (AI) techniques to partially optimize IDSs. For example, neural networks (NN), widely used for optimizing classification problems (Ripley, 1994; Golovko and Kochurko, 2005), have been also applied to improve misuse filtering and malicious pattern recognition (Lippmann and Cunningham, 2000; Lei and Ghorbani, 2004; Zhang et al., 2005). In (Ahmad et al., 2009), readers may find detailed survey on NN–based IDS. Moreover, Evolutionary Computation is especially suitable for those problems in which a cost–effort trade–off

exists such as event correlation (Suarez-Tangil et al., 2009). Thus, relative to the existing literature on improving SIEM systems by applying AI this position paper elaborate on the application of Artificial Immune Systems (AISs) to event correlation. Though it is not the first time this technique is considered in this domain, our approach is novel regarding the way event correlation is formulated.

1.2 Position Statement

As very promising solutions which are emerging by some sort of biological inspiration, AISs have been proven to contribute important benefits to different areas within computer security, since efficient abstractions of processes were found in the mid 1980s by (Farmer et al., 1986). In particular, several works focus on analyzing how immunological concepts may be applied to intrusion detection (Kim et al., 2007), pattern recognition and classification (Carter, 2000), anomaly detection, and distributed detection (Hofmeyr, 1999).

Perhaps the main advantage of AISs is that not only supervised learning is possible (Watkins et al., 2004), but also unsupervised (De Castro and Timmis, 2002; Timmis and Neal, 2001) indeed. In this position paper, we extend the typical architecture of a SIEM system to efficiently introduce an adaptive learning framework upon the correlation process.

To this regard, we position our paper with the following statements:

- Generally, the existing SIEM tools present limitations and contextual constrains. In addition, current SIEM frameworks deploy their own architecture. We propose a global framework which integrates the most promising research advances and formalizes an unified architecture design towards an intelligent correlation system.
- The strategy of combining intelligence and self–adaptation to optimize different types of computing services is emerging as a robust and efficient approach. In particular, we introduce a bio-inspired and adaptive learning based on AIS to enhance the SIEM as a whole but especially focusing on event correlation. Therefore, AIS-based SIEM systems will facilitate an adaptive correlation of novel multi–step attacks.

The rest of the paper is organized as follows. Section 2 describes the foundations of our work–in–progress. In Section 3 we establish the AIS entities needed to introduce an adaptive learning component into a traditional event correlation engine. Section 4 outlines the main phases of the proposed AIS-based event correlation framework. Finally, in Section 5 we establish the main conclusions as well as the immediate future work.

2 AN ARTIFICIALLY IMMUNE SIEM ARCHITECTURE

We informally introduce a novel approach based on artificial immune network[1] theory to both improve and extend traditional SIEM systems.

In particular, Artificial Immune Systems (AIS) extract and apply several interesting properties and concepts of the human immune system to provide solution to different types of computer process such as networks' defense against malicious actions. In fact, phenomena suggested by the biological adaptive immune response, against the encountered pathogens, metaphorically share common implications with process of attack pattern recognition.

Hence, we envision the construction of a SIEM system using an artificial immune network model. Our three-layer architecture (Fig. 1) comprising the following blocks:

- The physical barrier offers a physical protection against pathogens attacking the system as some sort of prevention layer.
- The Innate Immune System (present at birth in humans) deploys immune agents which are in charge of protecting the system against invaders as well as providing pattern recognition mechanisms.
- The Adaptive Immune System defines the logic to learn, adapt and memorize antigens and secrete the appropriate anti-body (i.e. correlation rules).

The complex adaptive system is the main focus of interest here that involves diverse and multiple interconnected elements, which tend to provide capacity to change and learn from experience. Several works have partially applied AIS to specific domains. An interesting approach for mapping the immunity entities and process on to the development of computational models is presented in (Dasgupta, 2006). In the following sections we further elaborate on the essential concepts which leads to an AIS-based implementation, namely:

1. The application domain: We must first define the main assumptions and definitions within this

[1]The immune network theory was first introduced by Jerne (Jerne, 1974) as a way to explain the memory and learning capabilities exhibited by the immune system. This theory has inspired a subfield of optimization algorithms as many other fields unrelated to biological immunology.

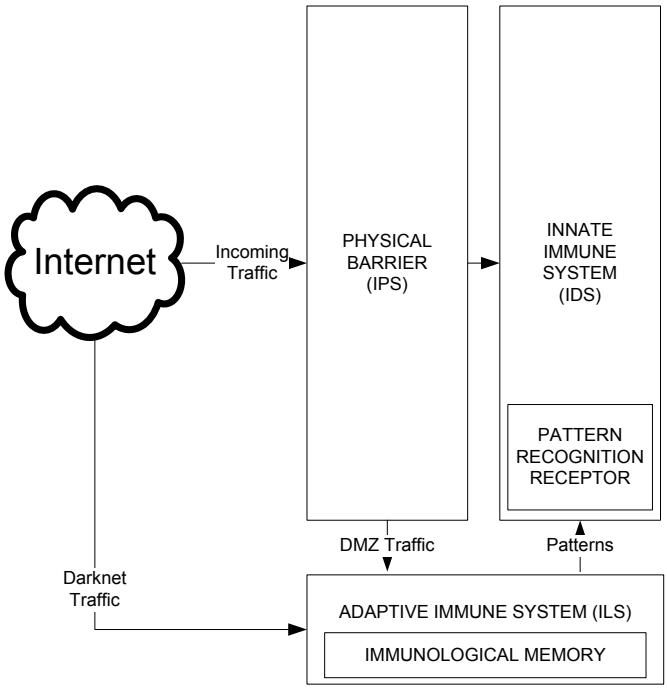

Figure 1: An event correlation framework based on AIS.

particular application domain and the correlation problem to be solved.

2. The immunity-based approach: As we will describe below, there are different techniques presented so far. To identify the most suitable AIS technique is not completely trivial in all instances.
3. Representation: We must establish the interpreted codification in terms of immune entities for the elements involved in the correlation context.
4. Adaptive immune algorithm: Finally, we define the immune algorithm aiming to automatically generate correlation fingerprints (i.e. most in the way of proteins).

2.1 Application Domain

We should consider the event correlation process from two different viewpoints: (i) which automatically learns without human supervision (Def. 2.1), and (ii) which requires an expert supervision (Def. 2.2):

Definition 2.1 (Automatic Supervision) *The process to extract the knowledge related to a novel attack without human intervention. Each extraction will form temporal correlation rules.*

Definition 2.2 (Expert Supervision) *An expert guides the process to extract the knowledge related to a novel attack and forms a permanent correlation rule.*

For the sake of completeness, we include the following basic concepts:

Definition 2.3 (Event) *Event is defined as a phenomenon produced when a particular security precondition becomes true. Generally such conditions denote established patterns extracted from previous interactions between two or more networking nodes.*

An event is defined by the features it consists of (namely its fingerprint):

Definition 2.4 (Event fingerprint) *The event's fingerprint consists of a set of attributes which identifies a certain event's properties. The specification of these attributes depends on the SIEM system.*

Event correlation aims at obtaining the fingerprint of a series of aggregated events. Thus,

Definition 2.5 (Event aggregation) *Event aggregation gathers together a collection of events which fulfill particular premises.*

Definition 2.6 (Event correlation) *Event correlation must probabilistically define the relationship between a set of aggregated events.*

and consequently,

Definition 2.7 (Correlation fingerprint) *A correlation fingerprint represents a correlation produced as a response of the successful relationship between a group of attributes.*

To this regards, we have identified the following attributes that completely characterize correlation fingerprints: *number of events* per each pair of sensor ID and Sub-ID; *protocols* per each pair of sensor ID and Sub-ID; number of *occurrences* per category; number of different *categories* per each pair of sensor ID and Sub-ID; total number of different *sources and destinations IP* addresses; total number of *events* and *categories*; maximum and minimum slot of *time*.

Note that it is out of the scope of this position paper to propose a complete taxonomy of event correlation features. We defer the discussion on these attributes to future work.

We position that intruder's actions swiftly evolve to become more effective, as well as more sophisticated generations of malware, i.e. polymorphic malware. In this regard, malware-analysis tools integrated along with an adaptive learning system will integrate our architecture to automatically generate specific correlation-fingerprints'.

Definition 2.8 (Darknet) *Also known as network telescope, is a system used to observe different large-scale events by monitoring unused network addresses.*

On the one hand, traffic ingoing to the darknet is, by definition, unrequested. We can assume that this traffic is likely to be generated by an intruder, and therefore can be labeled as malicious.

On the other hand, we can generate fake interactions with the intruder by emulating common exposed services. We assume here that despite aforementioned intrusion evasions, user's habits and used services will tend to behave alike. Thus, evolved malware will still interact with our emulated services. Honeynet (Def. 2.10) projects can be used in this regard (Spitzner, 2003).

Definition 2.9 (Honeynet) *A honeynet is a networked system used to trap malware by simulating to be an unprotected vulnerable resource, so that attacker's activities can be studied.*

In order to isolate each instance of a malware, we assume that we can forward traffic arriving to the darknet to an replica located in a DMZ.

Definition 2.10 (DMZ) *Demilitarized zone. Is defined as an area out of the boundaries of the network's organization, which is isolated from the protected assets.*

Additionally, we assume that also traffic detected as malicious by the SIEM can be redirected to a the DMZ for a better study of incoming anomalies.

2.2 Immunity-based Approach

There are several approaches defined in the literature to solve optimization problems using AIS (Kim et al., 2007) such as *dendric cell algorithm*, *gene libraries*, and *idiotypic networks*. The vast majority of the models proposed so far are bases on *immune network models* (Jerne, 1974), *clonal selection*, and *negative selection*.

AISs have been applied to different domains such as software fault prediction (Catal and Diri, 2009), and musical genre classification (Doraisamy and Golzari, 2010). The objective of these proposal focuses on producing efficient connections between the observed data and thus inferring an optimized conclusion. To the best of our knowledge, the AIS has not been relevant to the context of security information management until now.

Here, we briefly introduce the main immunity-based approaches which have translated some of the functions and behaviors of the mammalian adaptive immune system into bio-inspired computer processes:

I Clonal selection and Maturation: Generally used together, these models define the strategy to mitigate an infection, i.e. cloning the most successful antibodies (Kim and Bentley, 2001b). In particular, the maturation process introduces random variations over the antibodies cloned and thus increasing the probability to detect unknown behaviors.

II Negative selection: This technique is used after maturation phase aimed at identifying non–self cells from self-cells and deleting self-reacting cells (Kim and Bentley, 2001a). This algorithm is used for pattern recognition problems domains to obtain new patterns from available knowledge

Section 4 further elaborates on the aforementioned algorithms.

3 REPRESENTATION

In this section, we further examine all new concept related to the application of an AIS to the application domain described above.

Thus, a key principle within an AIS is now introduced, namely the proteins.

Definition 3.1 (Proteins) *Artificial immune theory defines the concept of secreting proteins as the mechanism used to detect non–self pathogens –malicious cells– which in turn are destroyed by antibodies. Proteins constitute the parameters to monitor and then distinguish self and non-self behaviors.*

Hence, two approaches for representing the application domain are possible: (i) when nodes represent the proteins role, and/or (ii) when events act as the subject to monitor. The former focuses on learning about the nodes which exhibit anomalous behavior. In this context, honest and misbehaving nodes embody self and non-self cells respectively. A major drawback of this approach is that compromised nodes generally produce both types of traffic, and therefore this could cause serious problems of false classification.

However, the alternative seems promising as identifying the events related to a certain attack present more similarities with traditional SIEM procedures. So, we formulate the following propositions as the basis of the artificial immune representation we propose (see Figure 2 for a quick look):

Claim 3.1 (Events) *Events are the AIS proteins. They will be classified as either self or non-self events.*

Claim 3.2 (Body) *The examined network depicts the biological body.*

Claim 3.3 (Self cells and Non–Self cells) *Authorized activity within the network will be classified as self by sensors. The opposite is cataloged as non–self.*

Claim 3.4 (Antigen) *An antigen is a series of events that matches an event fingerprint.*

Claim 3.5 (Antibody) *The antibody is the pattern (or rule) responsible of identifying a specific sequence of events.*

Claim 3.6 (Proteins) *Proteins are therefore the network activity monitored by sensors connecting to a central SIM. Proteins identify correlation fingerprints.*

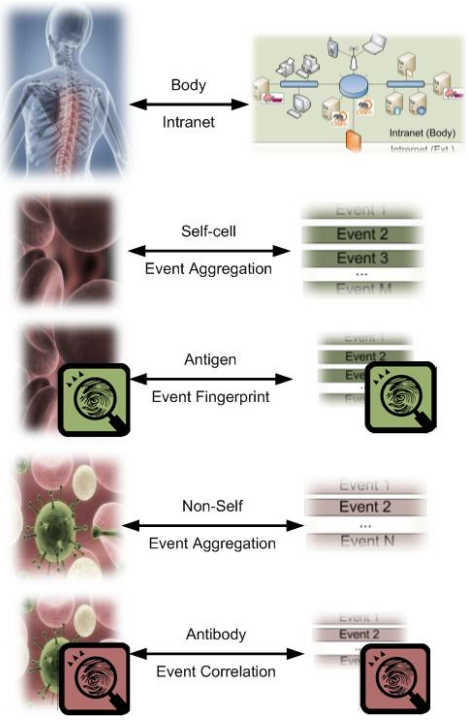

Figure 2: Mapping between immune and correlation entities.

4 IMMUNE ALGORITHM

In this section, we position the artificial immune algorithm as a three-phase protocol which combines traditional IDS concepts, data mining, honeynets and, just when strictly needed, the expert supervision, as follows:

I **Initial innate immune definition**. In this phase, the expert must define a range of values for the collection of correlation fingerprints. These values will act as discriminators for self cells. How to implement the appropriate value on the corresponding correlation fingerprint is critical. To this regard, existing repositories for known attacks and their associated correlation rules may be useful. Generally, this repository is named *Gene Library* and is essential for the process below.

II **Adaptive algorithm**. The adaptive algorithm produces a number of correlation fingerprints that will be used to learn new correlation rules. The basis of this algorithm relies on the AIS principles — random adaptations will produce new patterns by applying (i) antibody secretion, (ii) negative selection, (iii) pathogen match, and finally (iv) clonal selection based on affinity mutation.

III **Adaptive immunological memory consolidation**. New correlation fingerprints are consolidated based on the following knowledge extraction: using the expertise of the administrator and automation techniques. On the one hand, the expert has to manually inspect and validate the correlation rules in terms of their accuracy. On the other hand, honeynets seem the best candidate to assist the automated consolidation process. Specifically, generated fingerprints will be validated using the non-self activity reported on the darknet at the beginning. If any of the fingerprints matches then the immunological memory (associated to each correlation) will be increased.

5 CONCLUSIONS

In this position paper, we have discussed the application of AIS techniques to optimize current SIEM systems. To this regard, we propose an adaptive immune correlation system to be included into a typical SIEM architecture. Our global objective is to efficiently generate correlation rules and adaptively predict novel multi-step attacks. Our proposal comprises various strategies already used in intrusion detection, data mining, honeynets and, just when strictly needed, the expert supervision. Our hope is that this position paper will, directly or indirectly, inspire new directions on applying intelligence to security event correlation.

REFERENCES

Ahmad, I., Abdullah, A. B., and Alghamdi, A. S. (2009). Artificial neural network approaches to intrusion detection: a review. In *Proceedings of the 8th Wseas international conference on telecommunications and informatics*, pages 200–205. WSEAS.

Anuar, N., Papadaki, M., Furnell, S., and Clarke, N. (2010). An investigation and survey of response options for Intrusion Response Systems (IRSs). In *Information Security for South Africa (ISSA), 2010*, pages 1–8. IEEE.

Carter, J. H. (2000). The immune system as a model for pattern recognition and classification. *Journal of the American Medical Informatics Association: JAMIA*, 7(1):28–41.

Catal, C. and Diri, B. (2009). Investigating the effect of dataset size, metrics sets, and feature selection techniques on software fault prediction problem. *Information Sciences*, 179(8):1040–1058.

Dasgupta, D. (2006). Advances in artificial immune systems. *IEEE Comp. Intelligent Magazine*, 1(4):40–49.

De Castro, L. and Timmis, J. (2002). *Artificial immune systems: a new computational intelligence approach*. Springer Verlag.

Doraisamy, S. and Golzari, S. (2010). Automatic Musical Genre Classification and Artificial Immune Recognition System. *Advances in Music Information Retrieval*, page 391.

Farmer, J. D., Packard, N. H., and Perelson, A. S. (1986). The immune system, adaptation, and machine learning. *Physica D: Nonlinear Phenomena*, 22(1-3):187–204.

Golovko, V. and Kochurko, P. (2005). Intrusion recognition using neural networks. In *IEEE Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, pages 108–111. IDAACS.

Hofmeyr, S. (1999). *An immunological model of distributed detection and its application to computer security*. PhD thesis, University of New Mexico.

Jerne, N. K. (1974). Towards a network theory of the immune system. *Ann. Immunol.*, 125C:373–389.

Kim, J. and Bentley, P. (2001a). An evaluation of negative selection in an artificial immune system for network intrusion detection. In *Proc. of GECCO*, pages 1330–1337. Citeseer.

Kim, J. and Bentley, P. (2001b). Towards an artificial immune system for network intrusion detection: An investigation of clonal selection with a negative selection operator. In *Proc. of the 2001 Congress on Evolutionary Computation*, volume 2, pages 1244–1252. IEEE.

Kim, J., Bentley, P., Aickelin, U., Greensmith, J., Tedesco, G., and Twycross, J. (2007). Immune system approaches to intrusion detection–a review. *Natural computing*, 6(4):413–466.

Lei, J. Z. and Ghorbani, A. (2004). Network intrusion detection using an improved competitive learning neural network. In *Proc. of second annual conf. on communication networks and services research*, pages 190–197. IEEE Computer Society.

Lippmann, R. P. and Cunningham, R. K. (2000). Improving intrusion detection performance using keyword selection and neural networks. *Computer Networks*, 34(4):597–603. Recent Advances in Intrusion Detection Systems.

Mutz, D., Vigna, G., and Kemmerer, R. (2003). An Experience Developing an IDS Stimulator for the Black-Box Testing of Network Intrusion Detection Systems. In *Proc. of the 2003 Computer Security Applications Conf.*, Las Vegas, Nevada.

OSSIM (Visited July 2010). Open source security information management. `http://www.ossim.net/whatis.php` .

Ripley, B. (1994). Neural networks and related methods for classification. *Journal of the Royal Statistical Society*, 56(3):409–456.

Spitzner, L. (2003). The honeynet project: Trapping the hackers. *IEEE Security and Privacy*, pages 15–23.

Suarez-Tangil, G., Palomar, E., Fuentes, J. D., Blasco, J., and Ribagorda, A. (2009). Automatic rule generation based on genetic programming for event correlation. In *Computational Intelligence in Security for Information*, Advances in Soft Computing, pages 127–134, Burgos, Spain. Heidelberg, Springer Berlin.

Timmis, J. and Neal, M. (2001). A resource limited artificial immune system for data analysis. *Knowledge-Based Systems*, 14(3–4):121–130.

Watkins, A., Timmis, J., and Boggess, L. (2004). Artificial Immune Recognition System (AIRS): An Immune-Inspired Supervised Learning Algorithm. *Genetic Programming and Evolvable Machines*, 5(3):291–317.

Zhang, C., Jiang, J., and Kamel, M. (2005). Intrusion detection using hierarchical neural networks. *Pattern Recognition Letters*, 26(6):779–791.