ARTIFICIAL IMMUNITY-BASED CORRELATION SYSTEM

Guillermo Suarez-Tangil, Esther Palomar, Sergio Pastrana and Arturo Ribagorda
Department of Computer Science, University Carlos III of Madrid, Avda. Universidad 30, 28911 Madrid, Spain

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 semi-automatic generation of event correlation rules by applying Artificial Immune Systems (AISs).

1 PROBLEM STATEMENT

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, the continuous evolution of attacks, specially recent distributed, multi–step attacks, poses 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. 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).

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) have been also applied to improve misuse filtering and malicious pattern recognition (Lippmann and Cunningham, 2000). 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).

2 POSITION STATEMENT

As very promising solutions which are emerging by some sort of biological inspiration, Artificial Immune Systems[1] (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

[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.

Suarez-Tangil G., Palomar E., Pastrana S. and Ribagorda A..
ARTIFICIAL IMMUNITY-BASED CORRELATION SYSTEM.
DOI: 10.5220/0003610604220425
In *Proceedings of the International Conference on Security and Cryptography* (SECRYPT-2011), pages 422-425
ISBN: 978-989-8425-71-3

classification (Carter, 2000), to name a few. 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) 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, considering 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 a 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. Therefore, AIS-based SIEM systems will facilitate an adaptive correlation of novel multi–step attacks.
- Authors in (Hofmeyr and Forrest, 2000) propose LISYS, a framework for applying AIS to network security. Our approach presents two main differences. On the one hand, LISYS applies an AIS architecture to Network Intrusion Detection System (NIDS) whereas our application domain considers a centralized SIEM. Specifically, our approach could receive inputs from LISYS as a sensor to correlate complex multi-step attacks. On the other hand, our approach adds honeynets in order to minimize human intervention as far as possible. To the best of our knowledge, the AIS has not been relevant to the context of security event correlation by applying SIEM.

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

3 AN ARTIFICIALLY IMMUNE SIEM ARCHITECTURE

We envision the construction of a SIEM system using an artificial immune network model as a three-layer architecture (Fig. 1) comprising the following building blocks:

- The physical barrier offers a physical protection against pathogens attacking the system as some sort of prevention layer.

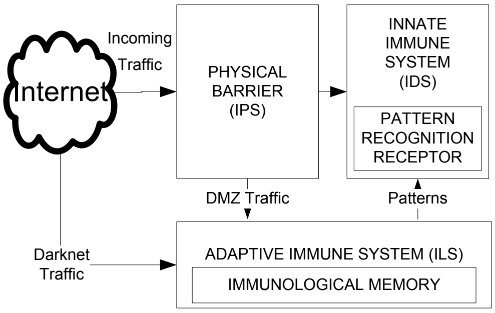

Figure 1: An event correlation framework based on AIS.

- 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.

3.1 Application Domain

We should consider the event correlation process from two different viewpoints:

Definition 3.1 (Automatic Supervision)**.** *The process to extract the knowledge related to a novel attack without human intervention. Each extraction will form temporal correlation rules.*

Definition 3.2 (Expert Supervision)**.** *An expert guides the process to extract knowledge related to a novel attack and forms a permanent correlation rule.*

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

Definition 3.3 (Event Fingerprint)**.** *Set of attributes which identifies certain event's properties. The specification of these attrib. depends on the SIEM system.*

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

Definition 3.4 (Event Aggregation)**.** *Event aggregation gathers together a collection of events which fulfill particular premises.*

Definition 3.5 (Event Correlation)**.** *Probabilistically define the relationship between a set of aggregated events. And consequently,*

Definition 3.6 (Correlation Fingerprint)**.** *Represents a correlation produced as a response of the successful relationship between a set of attributes.*

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

Therefore, 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 projects can be used in this regard (Spitzner, 2003).

3.2 Representation

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

Definition 3.8 (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.*

Now, 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.

We formulate the bases of the artificial immune representation (see Fig. 2) as follows. The examined network depicts the biological **body** and the events produced on the AIS will be represented as proteins. **Events** can be classified as either **self or non-self** events. Thus, authorized activity will be classified as self by sensors and the opposite is cataloged as non–self. Additionally, on the one hand, an antigen is a series of events that matches an event fingerprint. And on the other hand the antibody is the pattern (or rule) responsible of identifying a specific sequence of events. Therefore,

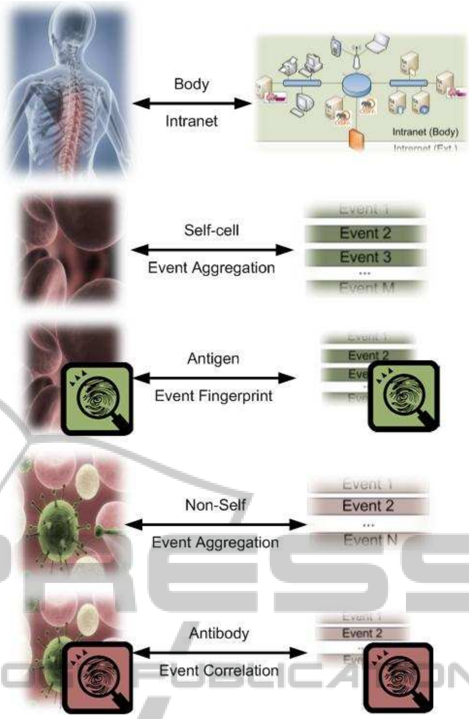

Figure 2: Mapping between immune and correlation entities.

Claim 3.1 (Proteins)**.** *Proteins are the network activity monitored by sensors connecting to a central SIEM. Proteins identify correlation fingerprints.*

3.3 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 (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.

4 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

Anuar, N., Papadaki, M., Furnell, S., and Clarke, N. (2010). An investigation and survey of response options for Intrusion Response Systems. 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.

Dasgupta, D. (2006). Advances in ais. *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.

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

Hofmeyr, S. A. and Forrest, S. (2000). Architecture for an artificial immune system. *Evolutionary computation*, 8(4):443–73.

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

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.

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 IDS.

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.

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.

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