TAT-NIDS: An Immune-Based Anomaly Detection Architecture for Network Intrusion Detection

Mário Antunes[1] and Manuel Correia[2]

[1] School of Technology and Management - Polytechnic Institute of Leiria, Alto do Vieiro, 2411-901 Leiria, Portugal
mario.antunes@estg.ipleiria.pt
[2] Faculty of Sciences - University of Porto, Rua do Campo Alegre 1021/1055, 4169-007 Porto, Portugal
mcc@dcc.fc.up.pt

Summary. One emergent, widely used metaphor and rich source of inspiration for computer security has been the vertebrate Immune System (IS). This is mainly due to its intrinsic nature of having to constantly protect the body against harm inflicted by external (*non-self*) harmful entities. The bridge between metaphor and the reality of new practical systems for anomaly detection is cemented by recent biological advancements and new proposed theories on the dynamics of immune cells by the field of theoretical immunology. In this paper we present a work in progress research on the deployment of an immune-inspired architecture, based on Grossman's Tunable Activation Threshold (TAT) hypothesis, for temporal anomaly detection, where there is a strict temporal ordering on the data, such as network intrusion detection. We start by briefly describing the overall architecture. Then, we present some preliminary results obtained in a production network. Finally, we conclude by presenting the main lines of research we intend to pursue in the near future.

Keywords: artificial immune system, tunable activation threshold, network intrusion detection, anomaly detection.

1 Introduction

The vertebrate Immune System (IS) [3] is an appealing metaphor and a very rich source of inspiration for new ideas on anomaly detection applied to network intrusion detection. It possesses two main layers of defense, termed *innate* and *adaptive*, whose main functions are to actively protect the body from intrusions of *pathogens*. The *innate* part of the immune system only recognizes specific known intruders by their "*signatures*", and its behavior is more less the same for all normal individuals in a given species. To overcome this limitation, the *adaptive* immune system has the ability to deal with a much more specific recognition of pathogens and to behave adaptively in order to detect heretofore unseen forms of intrusion. The IS has a complex set of different cell types that interact with each other by the means of chemical messages exchanges. The Antigen Presenting Cell (APC) digests and destroys the microorganisms into small *peptides* and then presents them to *lymphocytes* (*T-cells* and *B-cells*). These cells have specific *receptors* that can *bind* with a certain degree of affinity to the peptides present on the surface of each APC. Depending on the affinity level with the pathogen and on their activation threshold, the cells can become activated, thus initiating an immune response.

J.M. Corchado et al. (Eds.): IWPACBB 2008, ASC 49, pp. 60–67, 2009.
springerlink.com

A Network Intrusion Detection System (NIDS) is an application that monitors a computer network and identifies actions that can compromise its integrity and availability. Its main goal is to positively identify possible occurrences of ongoing attacks and, at the same time, to not be misled by false alarms. NIDS are usually classified into two main types: anomaly, behavior based and misuse, signature based (such as snort-IDS [2]). Metaphorically speaking, the IS challenges are very similar to those faced by a NIDS. Their major goals are to analyze in real time, the iterations of the "system" with the environment and to distinguish what corresponds to legitimate activities (self) from those that manifests themselves as potentially harmful actions (non-self). The adaptive IS behavior can also be seen as an anomaly detector, able to distinguish self from non-self activities. It acquires a "self" profile activity and, during the individual lifetime, adjusts the meaning of self according to the changes occurred in the environment, as well as its level of responsiveness. In the context of network intrusion detection, the meaning of self also changes dynamically through time.

The IS self-non-self discrimination processes inspired the Artificial Immune System (AIS) [6] research community to develop immune-inspired NIDS. There are two main types of immune-based NIDS developed so far [10]: those derived from classical Burnet's Negative Selection (NS) [4] and those that take advantage of Matzinger's Danger Theory (DT) [11]. Forrest's seminal work [7] and Kim's experiments [9] are the most relevant works using NS. The use of DT in the context of NIDS is being developed by Aickelin and colleagues [1]. More recently it has been demonstrated [13, 14] these immunological approaches have some serious limitations. The central problem is that there are no prior expectations that they can deal well with change. From one side, NS posits that evolutionary adaptation should have *fixed* immune system cells activation thresholds to an optimal value to ensure efficient self-nonself discrimination. DT bases its activity in built-in *danger* detectors and should eventually fail to detect harmful intrusions that avoid manifesting any of those. These limitations motivated us to investigate the appropriateness of developing an anomaly detection NIDS based on a different view of the IS activity: the Grossman's Tunable Activation Threshold (TAT) hypothesis [8]. The central idea behind TAT is that each immune cell has a tunable activation threshold whose value reflects the recent history of interactions with the surrounding environment. The potentially autoimmune (self) lymphocytes, which are continuously exposed to body antigens, raise their activation threshold and become unresponsive. In contrast, lymphocytes that are not auto reactive and recognize external microorganisms, have low thresholds and are fully responsive upon infection. Thus, the "classification" made by TAT is dynamic and depends only on the present intensity and rate of change of these signals. In summary, TAT requires no prior "classification" of the signals as either "self" or "non-self", and it is expected to automatically adjust each one of the individual cell dynamics into the current environment. The relevance of our research is two-fold. On the one hand, we investigate whether TAT possesses adequate adaptive characteristics to make it suitable to the *non-biological* environment of a computer network. On the other hand, we expect that results obtained within this context will lead to a better understanding of the scope of the TAT hypothesis as a valid and more coherent explanation of the behavior observed in a real vertebrate IS.

2 The TAT Model

We adopted a minimal mathematical model of TAT for T-cells [5]. In this model, T-cell activation is controlled by two enzymes that respond to antigenic signals (S) delivered by the APC: Kinase (K) and Phosphatase (P). The TAT dynamics for a T-cell in two different situations is depicted in Figure 1.

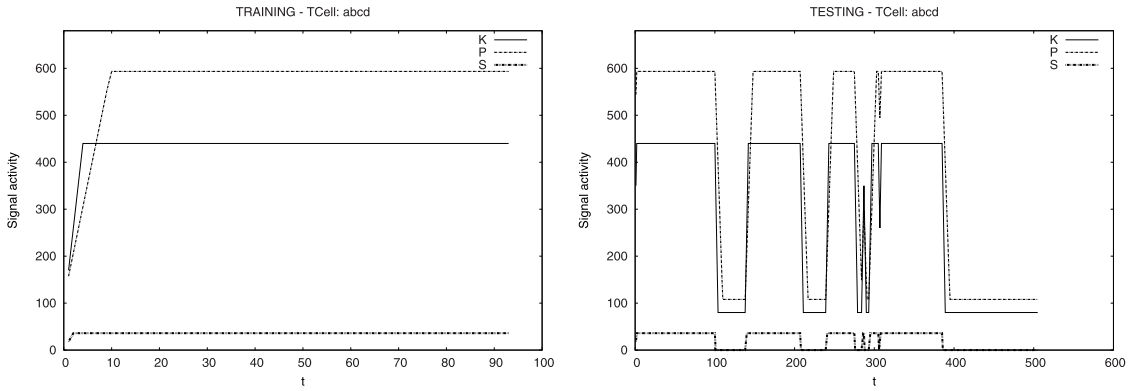

Fig. 1. TAT dynamics of two individual T-cells in the repertoire

In Figure 1(a) the T-cell receives a constant signal. Except for the initial transient, the condition $P > K$ is fulfilled for the remaining period, meaning that the T-cell will remain inactive. In Figure 1(b) the T-cell receives variable signals and adjusts K and P levels accordingly. During the transient periods in which the signaling is minimal, P and K levels tend towards the initial level, thus decreasing the activation threshold, in such a way that when the signaling increases again, K can transiently supersedes P, thus leading to repeated events of T-cell activation.

The T-cell dynamics exposed by these figures suggests the need to optimize some parameters in order to have good detection accuracy: the *slopes* of K and P (ϕK and ϕP), the maximum values for K and P (K_{max} and P_{max}), the minimum values for K and P (K_0 and P_0) and the affinity threshold between T-cells and peptides.

3 TAT-Based NIDS

The use of IS principles, models and components (such as cells and molecules) in problem solving, is framed into an AIS, whose basic architecture is generally composed by the following basic elements: *representation* of the immune system components, *evaluation* of its interaction within the environment and with each other, and *adaptation* of each component over time [6]. In this section we identify these key elements for an AIS based on TAT and describe the main building blocks for the developed TAT-based NIDS, depicted in Figure 2.

3.1 The Framework

The immune components represented in TAT-NIDS are the APCs, peptides and T-cells. The artificial APC, denoted APC, is a collection of *raw* network packets

("antigens") collected by the network interface (*Network traffic collector* module), encoded in *base64* and then concatenated together in a timely sequence with a predefined duration. This base64 APC stream is spliced into small sized strings ($PEPTIDES$), each corresponding to an artificial peptide. Each artificial cell, termed $TCELL$, is an object that receives signals from the peptides contained in APCs, compares them to a local string representing its unique specificity and adjusts its response threshold by tuning the values of its K and P variables. Table 1 summarizes the metaphor of IS terms in the context of the NIDS architecture proposed.

Table 1. The main immunological terms used in TAT model and its correspondence to our TAT-NIDS system

Immune System	**Network IDS**
Antigen Presenting Cell	APC
Peptide	$PEPTIDE$
MHC/Peptide complex	$PEPTIDE$ Pattern
T-Cell	$TCELL$
T-Cell Receptor (TCR)	$TCELL$ Pattern
T-Cell Repertoire	$TCELL$ Repertoire
Phagocytosis	Pre-processing
Specific recognition	Affinity ($TCELL$ and $PEPTIDE$)
Antigen	Packet
Autoimmune response	False positive
Self	"Normal" network packets
Non-self	possibly an ongoing attack

The core of the architecture is the *TAT-based AIS simulator*. It corresponds to a $TCELL$ dynamics simulator based on the TAT model. The APCs and $TCELL$s interactions and adaptation rate are calculated over time. The simulator processes two kinds of data-sets: a training and a testing data-set. The training data-set is further divided into two parts: a *calibration* and a *validation* part. Both have APCs with normal traffic but the later also has APCs classified as anomalous (Figure 2). In training mode the simulator loads the calibration part and interacts with a non-linear meta-heuristic simplex optimizer [12] that presents the simulator with a set of TAT parameters (Section 2) for minimizing the false alarms rate and simultaneously detecting the attacks included in the validation part. The main reason behind the introduction of attacks in training is to better guide the optimizer in finding a set of parameters that, not only minimizes the rate of false alarms, but can also achieve a low rate for false negatives for the network it has been trained to recognize. Otherwise, the parameters obtained are very likely to be too permissive and inefficient for the testing phase. Finally, in testing mode, the best parameters set obtained are then used to detect the anomalies present in the testing data-set.

The *artificial anomaly generator* module takes advantage of several network tools (such as `nmap`) to generate artificial network packets flows that match snort-IDS [2] rules or are caught by some snort-IDS preprocessing module.

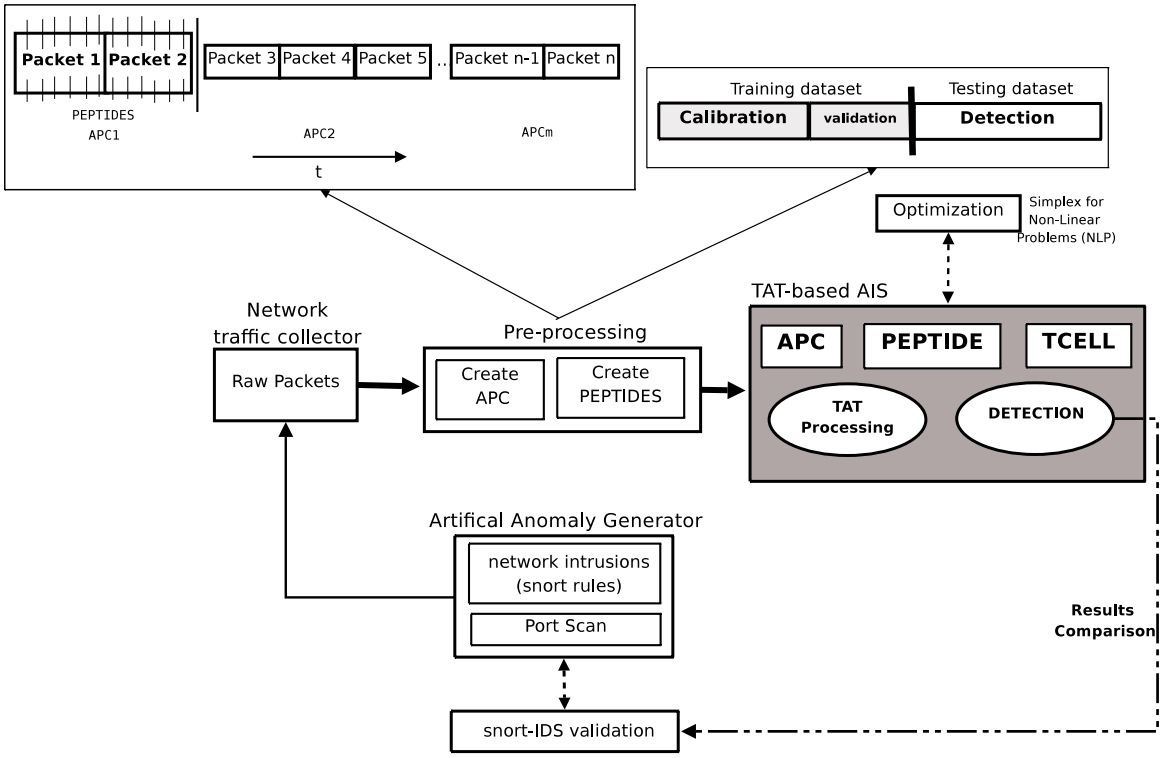

Fig. 2. General architecture of the TAT based NIDS

3.2 The Algorithm

The TAT based detection algorithm is depicted in Figure 3 and can be summarized as follow:

1. APCs are processed sequentially in a temporal basis.
2. Each APC presents its $PEPTIDE$s to the current repertoire of $TCELLS$. If no cell binds the $PEPTIDE$ with a strong affinity, a new one representing the

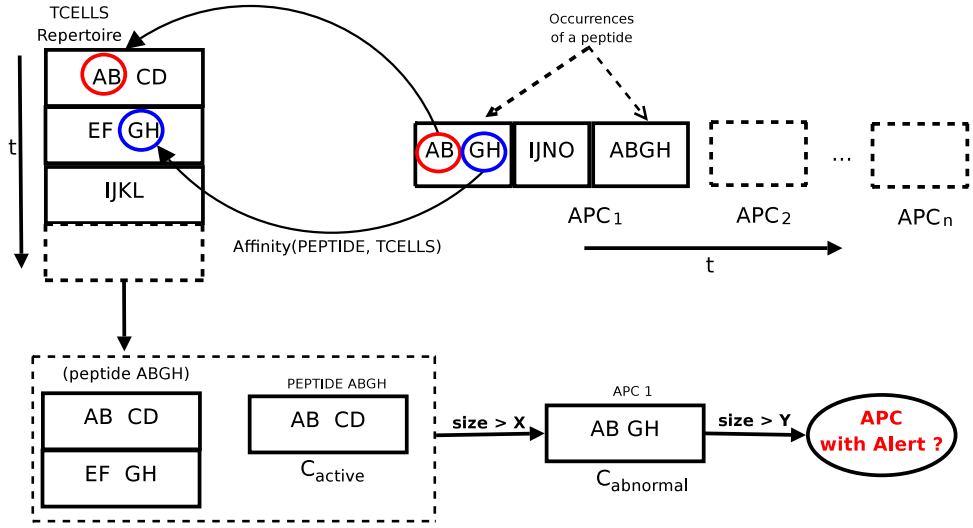

Fig. 3. The steps of the detection algorithm

$PEPTIDE$ is inserted into the repertoire. Otherwise, $TCELL$s that bind with the $PEPTIDE$ are stimulated with a signal (S).
3. According to TAT dynamics, some of those $TCELL$s will become *activated*, with $K > P$ (C_{active}).
4. If the ratio between the number of activated $TCELL$s and the total that bind with a $PEPTIDE$ is higher than a predefined threshold (X), then the $PEPTIDE$ is *classified* as abnormal ($C_{abnormal}$).
5. After processing all the $PEPTIDE$s in an APC, if the number of "abnormal" $PEPTIDE$s is higher than a percentage of the total number of $PEPTIDE$s in the APC (Y), then an alert is raised, indicating that the APC is probably abnormal (non-self).

The classification of an APC is then decided by the "committee" of $TCELL$s that become activated (with $K > P$). This is the role of the committees C_{active} and $C_{abnormal}$, that decide respectively on the classification of a particular peptide as *rare* and the whole of an APC as *abnormal*. These committee *thresholds*, for classifying peptides and APCs, are two parameters also optimized during the training phase.

4 Results

In this section we report preliminary results obtained by TAT-NIDS in two simple network experiments with intranet traffic. For the first we collected about 80 minutes of traffic that we have used for training (approximately 60 minutes for calibration and 20 minutes for validation) and 8 minutes for testing. For the test phase we have generated and inserted *four* attacks. For the simulator parameters optimization (see Section 3) we inserted *one* attack in the training data-set, different from those used for testing.

In the second experience we collected about 60 minutes of traffic that we used for training (approximately 48 minutes for calibration and 12 minutes for validation) and 19 minutes of test. During the test phase we have included an *IP Protocol Port Scan* sweep of the network, made by the *nmap* application. In both experiences, all the artificial attacks generated were detected by the snort-IDS.

With the set of optimized parameters presented in Table 2 we obtained the results described in Table 3. We were able to detect all the attacks (column Quantity) but with some False Positive (FP)s. The TAT algorithm could detect at least one of the attack packet-carrying APCs, which appeared contiguously in the testing data-set. This justifies the different values on column APCs when comparing the true positives and attacks. For instance, in the first experiment the detection of the fours attacks (in which the packets were distributed in 8 APCs), involved only the raising of an alert on 7

Table 2. TAT-NIDS optimized parameter set

Run	K0	P0	S0	Kmax	Pmax	ϕK	ϕP	$Affinity$ %	C_{active} %	$C_{abnormal}$ %
1	80.0	90.6424	8	10.0	11.3303	18	12.545	24.29	53.42	42.35
2	80.0	107.904	8	10.0	13.488	18	9.877	41.56	53.61	28.48

Table 3. Results obtained during the experiments

Run	Phase	APCs	$PEPTIDE$s	$TCELL$s	Attacks		True Positives		False Positives	
					Qty	APCs	Qty	APC	Qty	%
1	Training	916	4,244,899	63	1	2	1	1	5	0.5
	Testing	107	251,472	93	4	8	4	7	6	5.6
2	Training	726	6,387,471	77	1	2	1	1	8	1.1
	Testing	225	419,560	63	2	22	2	6	16	7.1

APCs. Thus, the detection was fully successful, even if not all the APCs were correctly classified.

5 Conclusions

In this paper we have presented an architecture for an AIS-NIDS based on the Grossman's TAT hypothesis. We have also reported some preliminary but promising results, obtained in a real network environment.

We have verified that TAT-NIDS is capable of detecting attack patterns present in the snort-IDS database, without having any prior information about them. These are however very preliminary results and we need to conduct much more extensive experiments with a much larger and richer data-set to be able to measure and better assess the real effectiveness of TAT in network intrusion detection. The major problems we have faced are mainly related to simulator parameter optimization. These happen because the TAT cell simulator parameters space set is extremely large and there is no apparent correlation between them.

We are currently comparing the TAT-NIDS with other AIS for anomaly detection, both in terms of performance and also in terms of accuracy. We are also testing our system on the detection of other vulnerabilities and exploits in the context of much more ambitious data-sets. We are also convinced that the framework we have developed so far can be easily applied, by modifying the preprocessing module that feeds the TAT simulator with APCs, to other complex domains where the objective is to detect some kind of "*temporal*" anomaly, like SPAM classification.

References

1. Aickelin, U., Bentley, P., Cayzer, S., Kim, J., McLeod, J.: Danger theory: The link between ais and ids? In: Timmis, J., Bentley, P.J., Hart, E. (eds.) ICARIS 2003. LNCS, vol. 2787, pp. 147–155. Springer, Heidelberg (2003)
2. Beale, J., Caswell, B.: Snort 2.1 Intrusion Detection. Syngress (2004)
3. Burmester, G.R., Pezzuto, A.: Color Atlas of Immunology. Thieme Medical Publishers (2003)
4. Burnet, F.M.: The Clonal Selection Theory of Acquired Immunity. Vanderbilt University Press (1959)
5. Carneiro, J., Paixão, T., Milutinovic, D., Sousa, J., Leon, K., Gardner, R., Faro, J.: Immunological self-tolerance: Lessons from mathematical modeling. Journal of Computational and Applied Mathematics 184(1), 77–100 (2005)

6. de Castro, L.N., Timmis, J.: Artificial Immune Systems: A New Computational Intelligence Approach. Springer, Heidelberg (2002)
7. Forrest, S., Perelson, A.S., Allen, L., Cherukuri, R.: Self-nonself discrimination in a computer. In: Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy, pp. 201–212 (1994)
8. Grossman, Z., Singer, A.: Tuning of activation thresholds explains flexibility in the selection and development of t cells in the thymus (1996)
9. Kim, J., Bentley, P.: An evaluation of negative selection in an artificial immune system for network intrusion detection. In: Genetic and Evolutionary Computation Conference 2001, pp. 1330–1337 (2001)
10. Kim, J., Bentley, P., Aickelin, U., Greensmith, J., Tedesco, G., Twycross, J.: Immune system approaches to intrusion detection - a review. Natural computing (2007)
11. Matzinger, P.: The Danger Model: A Renewed Sense of Self. Science's STKE 296(5566), 301–305 (2002)
12. Pedroso, J.P.: Simple Metaheuristics Using the Simplex Algorithm for Non-linear Programming. In: Stützle, T., Birattari, M., H. Hoos, H. (eds.) SLS 2007. LNCS, vol. 4638, p. 217. Springer, Heidelberg (2007)
13. Stibor, T., Timmis, J., Eckert, C.: On the appropriateness of negative selection defined over hamming shape-space as a network intrusion detection system. The 2005 IEEE Congress on Evolutionary Computation 2 (2005)
14. Vance, R.E.: Cutting edge commentary: A copernican revolution? doubts about the dangertheory. The Journal of Immunology 165, 1725–1728 (2000)