

Temporal Anomaly Detection: an artificial immune approach based on T-Cell activation, clonal size regulation and homeostasis

Mário J. Antunes ^{*}, Manuel E. Correia [†]

^{*}Polytechnic Institute of Leiria - School of Technology and Management
Leiria - Portugal

Email: mario.antunes@estg.ipleiria.pt

[†]CRACS/INESCN, DCC/FCUP, Oporto University
Porto - Portugal

Email: mcc@dcc.fc.up.pt

Abstract

The most popular anomaly detection systems use pattern matching mechanisms to search through a database of known anomaly signatures. These systems can only detect previously known anomalies, being thus unable to cope with new ones. There exists other popular kind of anomaly detection systems based on statistical approaches that compute normal behavior profiles. Moreover, since anomalies correspond to environment activities that deviate sufficiently from the pre-computed normality profiles, behavior based systems can generally cope with novelty anomalies. Unfortunately these systems suffer from a very high rate of false positives and do not cope well with gradual dynamic variations of normal behavior throughout time.

This paper presents an Artificial Immune System (AIS) based on Grossman's Tunable Activation Threshold (TAT) for temporal anomaly detection. We describe the immunological metaphor, the AIS components and the algorithm adopted for simulating T-cells behaviour, emphasizing two novel important features: the temporal dynamic adjustment of T-cells clonal size and its associated homeostasis mechanism. We present some promising results obtained with artificially generated data sets, aiming to test the appropriateness of using TAT in dynamic changing environments, to distinguish new unseen patterns as part of what should be detected as normal or as anomalous. We conclude by discussing results obtained thus far with these experiments.

1. INTRODUCTION

The natural Vertebrate Immune System (IS) evolved to become a highly complex defense mechanism that has the ability to recognize foreign substances (pathogens) and to distinguish between those that correspond to the harmless (self) from those that are related to some form of intrusion (non-self). The IS [1] is composed by two main layers of defense: *innate* and *adaptive*. The *innate* layer only recognizes specific known substances and its behavior is

similar in all individuals of the same species. In contrast, the *adaptive* layer is apparently unique to each individual and is able to “learn” and to recognize new forms of abnormal pathogens that gradually change throughout time, thus providing a highly sophisticated adaptive form of pathogen recognition. The IS is also supported by a complex set of cellular structures. Among them, the Antigen Presenting Cell (APC) digests and converts pathogens into small *peptides* which are then presented to *T-cells*, a lymphocyte, through a molecular structure denominated “MCH/Peptide Complex”. T-cells have a specific set of *receptors* that *binds* with a certain degree of affinity with the peptides that are being presented by APCs.

Nowadays, there exists already a full body of theoretical work involving models and algorithms devised by theoretical immunologists that describe and successfully predict certain aspects of the IS behaviour. This constitute the basis and source of inspiration behind the developments of Artificial Immune Systems (AIS) [2], [3], [4] for anomaly detection. AIS have already been deployed with some degree of success to network intrusion, computer virus and spam detection [5], [6], [7], among other applications.

AISs are usually divided into two major groups [7]. The first group comprises all AIS based on the classical Burnet's Negative Selection (NS) theory [8], like the Forrest's seminal work [9] and Kim's research [6] among others. The other group includes those that take inspiration on Matzinger's Danger Theory (DT) [10], comprising Aickelin's research the most well known [7], [11]. However, these systems suffer from some well documented serious limitations [12] [13]. This motivated us to investigate the appropriateness of developing an AIS based on a rather different biological theoretical perspective about the IS [14] [15]: the Grossman's Tunable Activation Threshold (TAT) hypothesis [16], [17]. TAT posits that each individual immune cell has its own 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, end up raising

their activation thresholds and thus become unresponsive. In contrast, lymphocytes that are not auto reactive and recognize external microorganisms, end up with low activation thresholds becoming thus fully responsive towards non-self antigens. TAT behaviour is thus completely dynamic. Its current state and emergent collective behaviour depends on the rate of change and intensity of the signaling for each cell that resulted from past interactions with the antigens each cell happens to get in contact with throughout time [16], [18]. In summary, for the AIS designer TAT assumes no prior "classification" of antigens as either "self" or "non-self", and it is expected to automatically adjust each one of the individual cell activation dynamics with the current environment, throughout time.

In our framework [14], we have also included two immunological concepts: cells clonal size regulation mechanisms and a dynamic equilibrium based on the sharing of finite resources (homeostasis). The former is related to the widely accepted idea from Biology that immune cells proliferate after being activated, cloning themselves and enabling a faster reaction for a second future encounter with that same pathogen [8]. This clonal selection process thus results in the growth of cell populations that became responsive, in order to trigger an appropriate immune response when in the presence of foreign and harmful pathogens. We mimic this concept by associating to each artificial cell a *weight* expressed by the number of clones of each T-cell. The later derives from the generally accepted idea that the immune cell repertoire is constantly changing by adapting to the current environmental circumstances. Since the amount of resources available within a body are finite and each immune cell competes with others for the same resources, the immune cell population as a whole cannot grow indefinitely, therefore each immune cell clone group should reach a state of dynamic equilibrium with the other groups. This process is denominated *homeostasis* and we have included it in our framework for dynamic clonal size adjustments. In our experiments we have been able to demonstrate that an AIS with appropriate homeostasis mechanisms possesses a better anomaly detection rate and can cope more gracefully with the presence of diverse stimulus.

In this paper we aim to describe a generic anomaly detection AIS based on the TAT theory (TAT-AIS) together with clonal size regulation and homeostasis mechanisms, as well as present results obtained with artificially generated data sets. In order to analyse how efficient our system really is on the dynamic toleration to unseen self patterns, we gradually replace throughout time the previously learned normal patterns by new unseen ones. Finally, we also analyse how efficient TAT-AIS is for anomaly detection by presenting the True Positive (TP) and False Positive (FP) rates obtained by the system with these experiments. The paper is organized as follows: in section 2 we explain in some detail the TAT concept and the model we have adopted for our AIS. In

section 3 we describe the TAT-AIS, its main components, features and describe the general algorithm behind TAT. In section 4 we describe the principles and methodology we have used to generate the artificial data sets and present some results obtained by running the TAT-AIS with these data-sets. In section 5 we discuss the results obtained, draw some conclusions and delineate guidelines for future research.

2. A MINIMAL TAT MODEL

The TAT theory hypothesizes that immune cell activation depends on a threshold that is adjusted dynamically and at each point in time it corresponds to the integrated received signalling past history. Every interaction between the cell receptor and the peptide ligands presented by the APC, results in an intracellular competition between "excitation" and "de-excitation" signaling pathways, causing the cell to adapt to the stimulus by increasing or decreasing its activation threshold. Therefore, cells with different antigen-specificity will have different activation thresholds as they are exposed to different stimuli. During its lifetime, each cell changes and adapts its responsiveness according to its interaction with the environment [16].

We have adopted a minimal mathematical model of TAT for T-cells [19]. Briefly, the model states that T-cell activation is controlled by two enzymes that respond to antigenic signals delivered by APCs: Kinase (K) and Phosphatase (P). Antigenic signals (S) lead to a linear increase of both K and P activities until they reach a plateau that is proportional to the intensity of the stimulus.

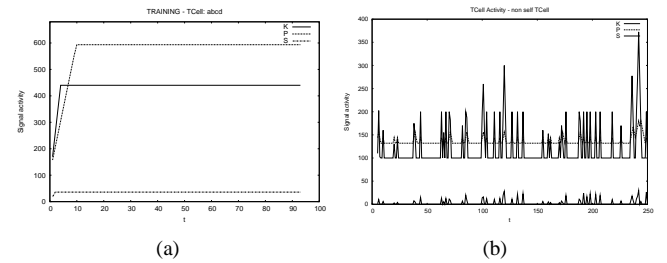


Figure 1: TAT dynamics of two individual T-cells in the AIS repertoire. 1(a): the T-cell receives a constant signal and becomes inactive ($P > K$) after an initial transient period. 1(b): the T-cell receives a variable signal and adjusts the K and P levels accordingly. This leads to repeated T-cell activation events where $K > P$.

For the same signal S , K increases faster than P , but if the signal persists P will eventually reach a higher plateau. Similarly, on signaling absence, K returns to the basal level at a faster rate than P . It is further assumed that T-cell activation is a switch-type response that requires that K supersedes P , at least transiently. Under these conditions, those T-cells that receive continuous or sufficiently frequent antigenic signals from APCs become unresponsive (Figure

1(a)) and those that rarely see their antigen remain sensitive [19] (Figure 1(b)).

Grossman and colleagues also postulated that T-cells that are tuned to be unresponsive by chronic exposure to a certain antigen could inhibit the activation of responsive T-cells in their neighborhood in physical and antigenic spaces [16]. This implies that an immune response will not depend on the sole response of an individual T-cell. Instead, it should happen if a certain group, a kind of *committee*, composed by the T-cells that became activated by a certain peptide interact in certain way before triggering an immune response [16], [19].

Within the immune system there are a finite number of cells, including T-cells. Clearly, an exponential growth in the number of T-cells is unsustainable in the medium term because there are limited resources available within the immune system to create and sustain a certain population of T-cells. This limitation on resources leads to a competition between T-cells cloning in order to guarantee that the repertoire size remains somewhat bounded in size. As a consequence, the repertoire has to reach some kind of equilibrium or *homeostasis*. It is thus expected that the clonal size of T-cells that recognize normal peptides remains in a steady state (homeostasis), sharing the total amount of resources available for all clonal T-cells populations existing within the whole repertoire. In practice this means that the clonal populations size of T-cells that recognize abnormal peptides should increase over time till a maximum homeostasis value is reached. This mechanism for growth and reduction for the clonal populations of activated T-cells thus conducts to a competition among the T-cells population “*strengths*” within the activated *committee* of active T-cells. We show with our experiments that this behaviour leads to a good detection rate for abnormal APCs

In our working model we define an *optimal committee size* and evaluate how accurate the detection gets with that value. We have also incorporated a mechanism of dynamic adjustment of clonal sizes, sustained by the use of homeostasis to control the weight each T-cell population has in the repertoire.

3. THE TAT-AIS FRAMEWORK

In the TAT-AIS, as with other AIS [2], there is a direct *mapping* of system components with the relevant biological IS immune counterparts. Table 1 summarizes the immunological metaphor adopted in our TAT model.

The *affinity* measure that we have adopted to represent the strength with which a certain T-cell binds with a peptide is calculated by the number of equal characters in the same positions on both strings. This function’s performance is very critical for TAT general processing efficiency, therefore it has to be very simple and very fast to compute.

Immune System	TAT-based AIS
Peptide	<i>PEPTIDE</i> (a string of characters).
MHC/Peptide ligand	<i>PEPTIDE</i> identification string.
Antigen Presenting Cell	<i>APC</i> . List of string (<i>PEPTIDE</i> s).
T-Cell	<i>TCELL</i> (artificial detector, identified by a string).
T-Cell Receptor (TCR)	<i>TCELL</i> identification string specificity
T-Cell Repertoire	List of available <i>TCELL</i> s.
T-cell Activation	<i>TCELL</i> value K is higher than P
Specific recognition	Distance between <i>TCELL</i> and <i>PEPTIDE</i> strings identifiers
Antigen	Contextual behaviour represented by a set of <i>APCs</i> .
Autoimmune response	False positive
Self	<i>TCELL</i> s that recognize recurrent (normal) patterns.
Non-self	<i>TCELL</i> s that recognize sporadic (abnormal) patterns.
Clonal size	<i>TCELL</i> attribute of trustworthiness used by decision <i>committee</i> .
Homeostasis	Weighted distribution of <i>TCELL</i> s clonal size in the admissible repertoire size.

Table 1: Immunological TAT metaphor.

In what follows we describe in some detail the TAT-AIS core framework as depicted by Figure 2 [14].

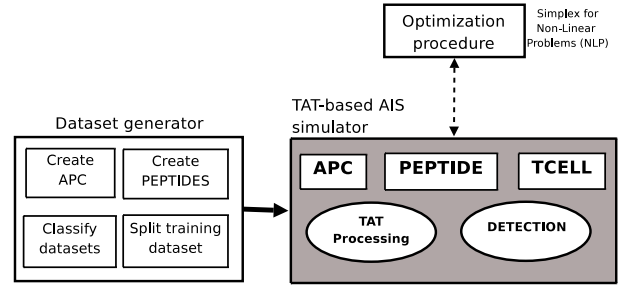


Figure 2: Building blocks of the TAT-based AIS.

The core of the architecture is the *TAT-based AIS simulator*. It corresponds to an artificial T-cell (*TCELL*) dynamics simulator based on the TAT model adopted. The *APCs* and *TCELL*s interactions and adaptation rate are calculated over time. The data set generator module pre-processes the antigens (usually an input file) in a format suitable for the simulator. During the validation part (see section 3.1) the simulator interacts with a non-linear meta-heuristic simplex optimizer [20] in order to obtain a set of parameters for the TAT simulator that simultaneously minimizes the false alarms rate and detects the known anomalies included on the corresponding training data set. The main reason behind the introduction of anomalies during 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 dataset it has been trained to recognize. Otherwise, the parameters obtained are very likely to be too permissive and inefficient for achieving an

accurate detection.

3.1. Experiments Methodology

Each of our experiments is divided in two distinct parts: *training* and *testing*, where each one has a corresponding data set (Figure 3). The training is further divided into two sub-phases: *validation* and “*vaccination*”. The validation phase aims to create an initial artificial T-cells repertoire that is able to recognize patterns that correspond to what is considered normal behavior.

During “vaccination” we present the system with a set of *APCs* whose artificial peptides (*PEPTIDES*) correspond to known anomalies. We process the “vaccination” data set with different TAT-AIS run parameter sets that are dynamically suggested by an optimizer (described in Section 3.3) in order to create an artificial T-cell (*TCELL*) repertoire that is able to recognize a set of patterns corresponding to a known abnormal behavior. This phase is essential to tune the system with a TAT-AIS parameter set that is able to produce good detection rates for anomalous patterns.

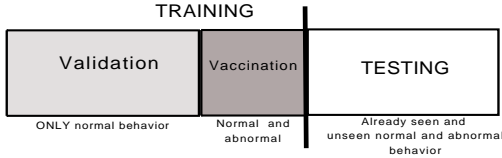


Figure 3: The methodology adopted for training and testing. Training is divided into validation and vaccination phases.

We then proceed to the testing phase where TAT-AIS is now confronted with a performance evaluation data set containing new unseen patterns to stimulate new or existing *TCELLs* (with different signal intensities). Throughout time, those *TCELLs* that bind with recurrent patterns should maintain $P > K$ and tolerate them as normal. Those *TCELLs* that bind with rarely seen patterns become much more responsive and become activated in the rare occasions where they appear.

3.2. Artificial data sets

The data sets used in the experiments have been artificially generated through a stochastic procedure. Each data set is comprised by *APCs* that represent a timely ordered set of events and the corresponding observed patterns. The core of each *APC* is composed by a list of strings separated by white space. These strings correspond to *PEPTIDES* that are being *presented* by the *APC*. For system performance evaluation purposes all *APCs* have been appropriately pre-tagged as “NORMAL” or “ALERT”.

An example of an *APC* (and its *PEPTIDES*) follows:

```
apc:764:NORMAL: cabc ccab gfhg (...) fhee efge
                  dcab dabb cadc dbdb cada
apc:765:ALERT:  aabc aBBB SPQP (...) BBAD bQQP
                  bbad 1332 RPPP CACA bPRQ
```

The data sets basic alphabet generators we have used for our experiments are shown in Table 2. All the *PEPTIDES* of a given artificial data set belong to the set of all possible fixed length strings (we have used length four) that can be obtained, allowing for character repetition, from the given alphabets.

Table 2: Alphabet used to create the *PEPTIDES*.

Alphabet	Data sets	APC Tag
{a,b,c,d}	Training, Testing	NORMAL
{a,A,B,C,D}	Training(vaccination), Testing	ABNORMAL
{b,P,Q,R,S}	Testing	ABNORMAL
{e,f,g,h}	Testing	NORMAL
{1,2,3,4}	Testing	NORMAL
{*,+,@,=}	Testing	NORMAL

We then use these *PEPTIDES* to randomly assemble an *APC*. If the *APC* has at least one abnormal *PEPTIDE* then it is tagged as “ALERT”, otherwise, the *APC* is tagged as “NORMAL”.

The *PEPTIDES* are randomly chosen from the list of admissible strings with a certain time *distribution*, as illustrated in Figure 4.

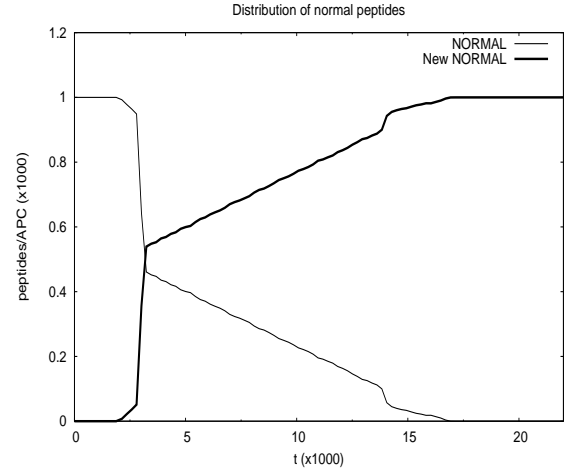


Figure 4: Distribution of normal occurrences during training and testing phases.

The normal behavior presented in the training dataset changes progressively in the testing phase, where normal *APC* content is gradually replaced by new unseen normal *PEPTIDES*. The *APCs* corresponding to unseen abnormal behavior appears sporadically in the testing dataset in a randomly temporal order. We have also included two different sets of normal *PEPTIDES* (“{1,2,3,4}” and

”{*,+,@,=}”) that only appear in the testing phase. These correspond to new normal events that occur through a certain period of time.

3.3. TAT-AIS run parameter set optimization

The TAT-AIS simulator requires some parameters to run. These are related to K and P dynamics, and are comprised by their initial values (K_0 and P_0), the slopes that define the way K (ϕK) and P (ϕP) change, the plateau values (K_{max} and P_{max}), the *affinity* threshold and a value for the detection threshold.

In order to reduce the number of simulation parameters and therefore to simplify their run time optimization, we have chosen to derive K_0 and P_0 and opted to fix all the parameters with the exception of K_{max} and ϕP , as described below:

$$\begin{aligned} K_0 &= S_0 * K_{max} & K_{max} &= \left(\frac{K_{max}}{P_{max}} \right) * P_{max} \\ P_0 &= S_0 * P_{max} & \phi P &= \left(\frac{\phi P}{\phi K} \right) * \phi K \end{aligned}$$

The ratios $\frac{K_{max}}{P_{max}}$ and $\frac{\phi P}{\phi K}$ varies between 0 and 1 and are optimised in the vaccination phase by an implementation of a meta-heuristic approach using a simplex algorithm for non-linear optimization [20]. In Section 4.2 we present the optimised parameters used for the experiments.

3.4. TCELL clonal size dynamics

The *TCELLs clonal size* is a number that represents the size of the sub-population of clones of that *TCELL*. Each time an *APC* is processed, the clonal size of affected *TCELLs* is updated as follows:

- 1) *TCELLs* are created with an initial clone size value (we adopted $C_0 = 2$).
- 2) **Validation phase:** For activated *TCELLs* ($K > P$), the clonal size (C) increases. Otherwise, C decreases.
- 3) **Vaccination phase:** For activated *TCELLs* in a tagged abnormal *APC* (a true positive), C increases. Otherwise C remains the same.
- 4) **Testing phase:** For *TCELLs* activated by an abnormal *APC*, C increases. For those bound but not activated ($P > K$), C increases but with a low rate. Otherwise, C decreases.

When an abnormal *APC* is detected, the activated *TCELLs* *proliferate* by increasing its clonal size. Otherwise, those *TCELLs* activated in the presence of a normal *APC* are subjected to a penalty, i.e. they decrease their clonal size. Those *TCELLs* with $P > K$ will participate in the *homeostasis* process, sharing the repertoire size of *TCELLs* that recurrently bind with “normal”

PEPTIDEs. We defined that C increases and decreases by units of two. For *TCELLs* with $P > K$ we fixed the increase in the order of 0.5 units. These were the constant values that gave us good results in practice.

3.5. TAT-AIS execution procedures

Given a data set comprised by a certain number of *APCs*, TAT-AIS execution proceeds as follows:

- 1) Each *APC* is processed sequentially, following the temporal order given by its identification number.
- 2) Each *APC* presents its *PEPTIDEs* to the current repertoire of *TCELLs*. For each *PEPTIDE*, if no cell binds with an affinity higher than a predefined threshold, a new cell representing the *PEPTIDE* string is inserted into the general repertoire. Otherwise, *TCELLs* that bind with the *PEPTIDE* are stimulated with a signal (S) that corresponds to the number of occurrences of the *PEPTIDE* in the *APC* times its affinity with the *TCELL*. Thus, an *APC* can stimulate several *TCELLs* with different signals intensities, depending on the number of occurrences and each individual cell affinity.
- 3) According to the TAT dynamics (see Section 3.6), some *TCELLs* will become *activated*, with $K > P$.
- 4) If the ratio between the sum of all clonal sizes of activated *TCELLs* and the sum of all clonal sizes of all the *TCELLs* that bind to the *PEPTIDEs* presented by the *APC* is higher than a predefined threshold, then the *APC* is “classified” as abnormal.
- 5) After processing all the *PEPTIDEs* in the *APC*, the clonal size is adjusted in all the participating *TCELLs*, according to the homeostasis algorithm described above (Section 3.4).

3.6. TCELL dynamics

Algorithm 1 depicts the *TCELLs* enzyme dynamics based on TAT, according to the signal received from the *APC* [19].

Algorithm 1 Update *TCELL* parameters.

```

1: if  $((S + S_0) * K_{max}) > K$  then
2:    $K \leftarrow MIN((S + S_0) * K_{max}, K + \phi K * \Delta t)$ 
3: else
4:    $K \leftarrow MAX((S + S_0) * K_{max}, K - \phi K * \Delta t)$ 
5: end if
6: if  $((S + S_0) * P_{max}) > P$  then
7:    $P \leftarrow MIN((S + S_0) * P_{max}, P + \phi P * \Delta t)$ 
8: else
9:    $P \leftarrow MAX((S + S_0) * P_{max}, P - \phi P * \Delta t)$ 
10: end if

```

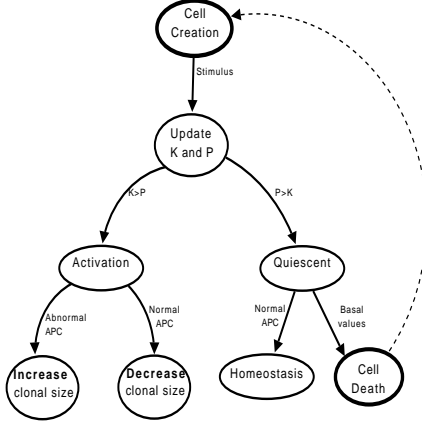


Figure 5: *TCELL* life cycle.

3.7. *TCELL* life cycle

Figure 5 depicts the *TCELL* life cycle. At first, a new *TCELL* is created when the pattern represented by the *PEPTIDE* does not match any one in the current repertoire. When there is a match with an appropriate affinity, each matching *TCELL* updates its K and P values according to the signal sent by the *APC* (Algorithm 1). If a *TCELL* becomes active, there are two possible outcomes: (1) the *APC* is abnormal and the *TCELL* will proliferate, increasing its clonal size; (2) the *APC* is normal and the *TCELL* decreases its clonal size. If the *TCELL* remains quiescent (with $P > K$) it will contribute to the *homeostasis* process. Finally, a *TCELL* will die if it is not stimulated for a certain predefined period of time, thus decreasing gradually its K and P values to its initial values (K_0 and P_0) and its clonal size is below a predefined minimum.

4. Evaluation and Results

Our working hypothesis is that TAT-AIS is able to recognize new unseen patterns and is able to further distinguish between those patterns that are considered self from others, included in *APCs* related to abnormal activities and identified as non-self. According to TAT, this self-non-self distinction is based on the temporal historic frequencies of patterns presented by past *APCs*. Through time, the detectors (*TCELLs*) that recognize frequent patterns become inactive and evolve to a quiescent state while those that detect sporadic patterns within *APCs* with a reasonable concentration of abnormal patterns, become reactive thus initiating an immune response. We have tested the system with artificial data sets specially constructed to include a gradual variation of what can be considered as normal behavior throughout time. In these data sets, previously unseen patterns start appearing more often and gradually replace the normal patterns learned during the training phase

(Figure 4 in Section 3). Our aim was thus to test the TAT detection algorithm capability to adapt to such a changing environment.

4.1. Experimental setup

We have conducted experiments with two different artificial data sets, as illustrated in Table 3. The affinity is measured by the distance between the cell receptor and ligand representative strings, calculated by the number of equal characters at the same position in the string. We fixed two different affinity measures for each run: 25% and 50%, corresponding to one and two equal characters respectively. The detection threshold, calculated by the ratio between bound and activated *TCELLs* clonal size for each *APC*, was also fixed at 0.2 (20%) for all the experiments. The vaccination data set comprises 25% of the whole training data set.

Table 3: Size of training, vaccination and testing data sets.

Run	Abnormal <i>APCs</i>		Total <i>APCs</i>		
	Vaccination	Testing	Training	Vaccination	Testing
1-2	150	46	1500	500	20000
3-4	150	47	1500	500	30000

4.2. TAT Parameters

The TAT parameters have been fixed with the values presented in Table 4. H_{max} corresponds to the maximum repertoire size of normal *TCELLs* for homeostasis. C_{max} is the maximum clonal size reached by an individual *TCELL* that tends to recognise abnormal patterns and is calculated by $C_{max} = H_{max} * 0.2$.

Table 4: Fixed parameters

S_0	C_0	K_{max}	ϕK	C_{max}	H_{max}	Threshold
10	2	10	10	51.2	256	0.2

We then employ the parameter set found by the optimizer that achieved the best detection rate in the “vaccination” phase for the testing phase (Table 5). In this case, the best detection rate corresponds simultaneously to the highest number of true positives and the lowest number of false positives.

Table 5: Optimized parameters obtained for the vaccination phase.

Run	$\frac{K_{max}}{P_{max}}$	$\frac{\phi P}{\phi K}$	<i>Affinity</i>
1	1.356370	0.473537	0.25
2	1.348210	0.138334	0.50
3	1.450690	0.460546	0.25
4	1.438480	0.484561	0.50

4.3. Results

Table 6 illustrates the results obtained for each experiment. The first three columns contain the repertoire size for the end of each execution phase. The columns TP and FP represents the quantity and percentage of both true and false positives obtained for *testing*.

Table 6: Results obtained during experiments.

Run	Repertoire size			TPs		FPs	
	Training	Vaccination	Testing	Qty	%	Qty	%
1	135	194	255	46	100	180	0.9
2	135	258	345	46	100	394	1.9
3	212	295	454	47	100	368	1.2
4	212	380	504	47	100	220	0.7

In spite of the great number of patterns that the system is exposed to, TAT-AIS possesses an efficient cell death mechanism (apoptosis) that is capable of maintaining an effective low cell repertoire size. It is also worth noting that even with a total replacement of the normal behavior during the testing phase, the TAT-AIS obtained a full detection rate with a relatively low number of false positives (less than 2%) for all the experiments.

During the *testing* phase, *TCELLs* clonal size changes according to whether the *TCELLs* recognizes normal or abnormal patterns, as depicted in Figure 6. For example the *TCELL* ‘abcd’ binds only with normal patterns and its clonal size contributes to the system *homeostasis* equilibrium. In another direction, *TCELL* ‘PSbR’ binds sporadically to the *PEPTIDES* presented by abnormal *APCs*, thus its clonal size increases throughout time.

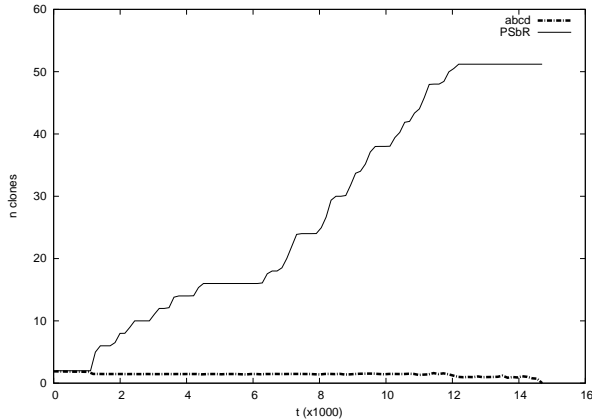


Figure 6: Clonal size dynamics.

The signalling activity of both *TCELLs* is depicted in Figure 7. For a normal *TCELL* like ‘abcd’, after an initial period of time, the ratio $\frac{K}{P}$ is below one, which means the *TCELL* is inactive (7(a)). Otherwise, for those associated with an abnormal pattern like ‘PSbR’, the ratio is transiently below and above one, reflecting successive activations (7(b)).

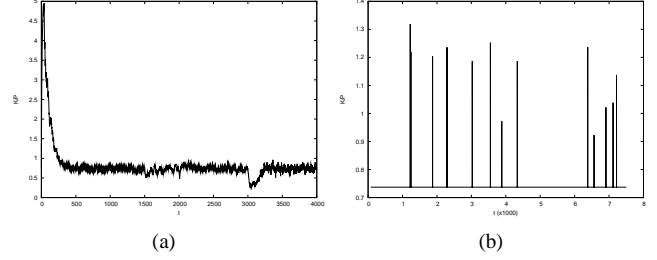


Figure 7: *TCELLs* signal activity. 7(a): *TCELL* with $TCR = abcd$ recognizes a normal pattern. 7(b): *TCELL* with $TCR = PSbR$ recognizes abnormal *PEPTIDES* in the *APCs*.

5. DISCUSSION

We have presented TAT-AIS, a generic TAT-based AIS for temporal anomaly detection and described its main architectural components. We have also presented some results obtained with two artificially generated data sets of predefined patterns resulting from normal and abnormal behaviors. These results are in line and improved upon with some previously published results about TAT-AIS [14], [15], [21]. Our aim was to validate the appropriateness of using TAT to detect new previously unseen patterns and also to distinguish them between those that correspond to unseen ‘normal’ and ‘abnormal’ behaviors.

We have observed that TAT-AIS has interesting properties for anomaly detection, provided the following basic generic requirements is true: *normal behavior is frequent and abnormal behavior is sporadic in time*. By *frequent* we mean a pattern that repeatedly stimulates a set of *TCELLs* that through time, by the TAT dynamics, stabilizes its enzymatic values ($P > K$). On the other hand, by *sporadic* we mean a pattern that stimulates intermittently a set of *TCELLs* with such a signal that implies its activation ($K > P$). Through time, these *TCELLs* increase gradually its clonal size and become more reactive to further similar recognitions. We have also emphasized the importance of having a ‘vaccination’ phase that presents *APCs* with known anomalies and fine tune the run time parameters set that is going to be used for the testing evaluation phase.

In TAT-AIS, detection is also dependent on the *TCELLs* dynamical clonal size control mechanism. We have presented an heuristic and have defined rules to adjust each *TCELLs* clonal size. We have adopted an homeostasis mechanism for *TCELLs* that helps to recognize self patterns, allowing the abnormal *TCELLs* to grow up until a predefined plateau is reached. This execution model reinforces the importance of having T-cell ‘population committee’ decisions based on *clonal size* dynamics.

The results thus obtained with TAT-AIS are very satisfactory, achieving a high rate of detection and a low level of false positives on the stochastic data sets we have produced.

It is worth to emphasize that in these data sets, during the testing phase, normal behavior is made to change gradually throughout time. In spite of this, TAT-AIS is able to correctly detect the new patterns and to correctly sort them into normal and anomalous sets. Moreover, based on these empirically results, we believe that TAT-AIS can compete with other approaches on the self-non-self distinction for dynamic environments that tend to change gradually throughout time their normality behavior profile. Firstly, the clonal size and activation threshold of each *TCELL* is being continuously updated according to its interactions with the environment. This means that each *TCELL* keeps track of its historical activity and, through time, delineates a trend to detect patterns corresponding to normal or abnormal behaviors. Secondly, by taking advantage of the “vaccination” phase, TAT-AIS can learn patterns of known anomalies that may improve the detection of both known and unknown anomalous behaviors during the testing phase.

We are well aware that these stochastic data sets were artificially generated and are most certainly not completely representative of real world phenomenons like the data sets we could obtain for example by live network traffic collection or computer systems trace logs. We have however already obtained some preliminary good results with a simpler less sophisticated TAT execution model, applied to network intrusions detection with real network traffic [14]. We are also confident that the current more sophisticated TAT-AIS will uphold and improve upon those same good results.

References

- [1] G. Burmester and A. Pezzuto, *Color Atlas of Immunology*. Thieme Medical Publishers, 2003.
- [2] L. de Castro and J. Timmis, *Artificial Immune Systems: A New Computational Intelligence Approach*. Springer, 2002.
- [3] D. Dasgupta, “Advances in artificial immune systems,” *IEEE Computational Intelligence Magazine*, vol. 1, no. 4, pp. 40–49, 2006.
- [4] L. Castro and J. Timmis, “Artificial immune systems as a novel soft computing paradigm,” *Soft Computing-A Fusion of Foundations, Methodologies and Applications*, vol. 7, no. 8, pp. 526–544, 2003.
- [5] S. Hofmeyr and S. Forrest, “Immunity by design: An artificial immune system,” *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pp. 1289–1296, 1999.
- [6] J. Kim and P. Bentley, “An evaluation of negative selection in an artificial immune system for network intrusion detection,” *Genetic and Evolutionary Computation Conference*, pp. 1330–1337, 2001.
- [7] J. Kim, P. Bentley, U. Aickelin, J. Greensmith, G. Tedesco, and J. Twycross, “Immune system approaches to intrusion detection - a review,” *Natural Computing*, vol. 6, no. 4, pp. 413–466, 2007.
- [8] F. Burnet, *The Clonal Selection Theory of Acquired Immunity*. Vanderbilt University Press Nashville, Tenn, 1959.
- [9] S. Forrest, A. Perelson, L. Allen, and R. Cherukuri, “Self-nonsel discrimination in a computer,” *Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy*, pp. 201–212, 1994.
- [10] P. Matzinger, “The Danger Model: A Renewed Sense of Self,” *Science’s STKE*, vol. 296, no. 5566, pp. 301–305, 2002.
- [11] U. Aickelin and S. Cayzer, “The danger theory and its application to artificial immune systems,” *proceedings of The First International Conference on Artificial Immune Systems (ICARIS 2002)*, pp. 141–148, 2002.
- [12] T. Stibor, J. Timmis, and C. Eckert, “On the appropriateness of negative selection defined over hamming shape-space as a network intrusion detection system,” *Evolutionary Computation, 2005. The 2005 IEEE Congress on*, vol. 2, 2005.
- [13] R. Vance, “Cutting edge commentary: A copernican revolution? doubts about the danger theory,” *The Journal of Immunology*, no. 165, pp. 1725–1728, 2000.
- [14] Authors, *Proceedings of IWPACBB’08 - Advances in Soft Computing (Springer)*, 2008.
- [15] —, *Proceedings of International Conference of Bio-inspired systems and signal processing (BIOSIGNALS 2009)*, 2009.
- [16] Z. Grossman and W. Paul, “Adaptive cellular interactions in the immune system: The tunable activation threshold and the significance of subthreshold responses,” *Proceedings of the National Academy of Sciences*, vol. 89, no. 21, pp. 10 365–10 369, 1992.
- [17] Z. Grossman and A. Singer, “Tuning of activation thresholds explains flexibility in the selection and development of T cells in the thymus,” *Proceedings of the National Academy of Sciences*, vol. 93, no. 25, p. 14747, 1996.
- [18] H. van den Berg and D. Rand, “Dynamics of T cell activation threshold tuning,” *Journal of Theoretical Biology*, vol. 228, no. 3, pp. 397–416, 2004.
- [19] J. Carneiro, T. Paixão, D. Milutinovic, J. Sousa, K. Leon, R. Gardner, and J. Faro, “Immunological self-tolerance: Lessons from mathematical modeling,” *Journal of Computational and Applied Mathematics*, vol. 184, no. 1, pp. 77–100, 2005.
- [20] J. Pedroso, “Simple Metaheuristics Using the Simplex Algorithm for Non-linear Programming,” *Engineering Stochastic Local Search Algorithms. Designing, Implementing and Analyzing Effective Heuristics - LNCS*, vol. 4638, pp. 217–221, 2007.
- [21] Authors, “abc,” *Proceedings of International Joint Conferences on Bioinformatics, Systems Biology and Intelligent Computing (IJCBS’09)*, pp. 1–1, 2009.