TOWARDS AN IMMUNE-INSPIRED TEMPORAL ANOMALY DETECTION ALGORITHM BASED ON TUNABLE ACTIVATION THRESHOLDS

Mário Antunes[1], Manuel Correia[2] and Jorge Carneiro[3]
[1] *School of Technology and Management, Polytechnic Institute of Leiria, Morro do Lena-Alto do Vieiro, Leiria, Portugal*

[2] *Department of Computer Science, University of Porto, Rua do Campo Alegre, 1021-1055, Porto, Portugal*

[3] *Instituto Gulbenkian de Ciência, Oeiras, Portugal*

Keywords: Artificial immune system, Anomaly detection, Tunable activation threshold, T-cell simulation and modelling, Pattern recognition.

Abstract: The detection of anomalies in computer environments, like network intrusion detection, computer virus or spam classification, is usually based on some form of pattern *search* on a database of *"signatures"* for known anomalies. Although very successful and widely deployed, these approaches are only able to cope with anomalous events that have already been seen. To cope with these weaknesses, the "behaviour" based systems has been deployed. Although conceptually more appealing, they have still an impractical high rate of false alarms. The vertebrate Immune System is an emergent and appealing metaphor for new ideas on anomaly detection, being already adopted some algorithms and theoretical theories in particular fields, such as network intrusion detection. In this paper we present a temporal anomaly detection architecture based on the Grossman's Tunable Activation Threshold (TAT) hypothesis. The basic idea is that the repertoire of immune cells is constantly tuned according to the cells temporal interactions with the environment and yet retains responsiveness to an open-ended set of abnormal events. We describe some preliminary work on the development of an anomaly detection algorithm derived from TAT and present the results obtained thus far using some synthetic data-sets.

1 INTRODUCTION

The vertebrate Immune System (IS) (Sompayrac, 2008) inspired the deployment of Artificial Immune Systems (AIS) (de Castro and Timmis, 2002) and has already been successfully used as a promising source of inspiration for new ideas on anomaly detection (Kim et al., 2007). The IS is an extremely complex distributed system whose main function is to actively protect the body from the intrusion of *pathogens*. It is composed by two main layers of defense: *innate* and *adaptive*. The *innate* part only recognizes specific known intruders by their "*signatures*", and its behavior is similar in all individuals of the same species. In contrast, the *adaptive* part is in a sense unique to each individual and is able to "learn" throughout time to recognize new forms of intrusive pathogens, thus providing a much more specific and adaptive form of recognition of pathogens.

The IS is supported by a complex set of cells. The Antigen Presenting Cell (APC) digests and converts pathogens into small *peptides* which are then presented to *T-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 and thus become activated.

Anomaly detection can be seen as a technique that produces a model for identifying cases that in some way deviate from a "learned" normal behavior. Decisions are based on a *profile* of normal behaviour and an anomaly is any particular case instance that is an *outlier* under this characterization. Current anomaly detection systems are mainly based on statistics, data-mining, data fusion and bio-inspired approaches, like neural networks.

Interestingly, the problem of creating a system capable of monitoring a normally changing environment and yet retaining the capacity to detect open-ended anomalies has been developed by natural selection during the evolution of the vertebrate IS. This sys-

Antunes M., Correia M. and Carneiro J. (2009).
TOWARDS AN IMMUNE-INSPIRED TEMPORAL ANOMALY DETECTION ALGORITHM BASED ON TUNABLE ACTIVATION THRESHOLDS.
In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing*, pages 357-362
DOI: 10.5220/0001553303570362

tem is capable of discriminate and engage in very different ways both *normal* body components and very similar but foreign (*abnormal*) chemical structures present in microorganisms. The IS is also able to learn and memorise the first encounter it has with these intruders, and can make effective use of this acquired knowledge to better deal with them on a future encounter. Perhaps even more relevant for the designing of an effective anomaly detection system, is the now well accepted fact that the IS learns the body composition during embryo life and adapts to physiological changes as the individual matures and ages (notable examples being hormones during sexual maturation or metamorphoses in some vertebrates).

Negative Selection (NS) (Forrest et al., 1994) and Danger Theory (DT) (Greensmith et al., 2006) were the immune theories most used on the development of IS-based anomaly detection systems (Kim et al., 2007). In this paper we explore a different view of the immune system to present the first developments of a new anomaly detection algorithm based on the Tunable Activation Threshold (TAT) hypotheses put forward by Grossman and colleagues (Grossman and Paul, 1992). In TAT it is assumed that lymphocytes have tunable activation thresholds whose value reflects the recent history of signaling they have been receiving from the environment. Potentially autoimmune lymphocytes, which are continuously exposed to body antigens raise their activation threshold, and become unresponsive or anergic. In contrast, lymphocytes that are not auto-reactive but recognise microorganism structures have low activation thresholds and are thus fully responsive upon infection.

This paper is organized as follows: in section 2 we explain in some detail the TAT concept and the model dynamics we have used for T-cells. In section 3 we describe the system architecture, its main components and features, as well as a methodology we have used for the generation of synthetic data-sets we are using to evaluate the system in a controlled way. In section 4 we present the results obtained with the experiences we have done with the artificially generated generic data-sets. In section 5 we discuss the results obtained, draw some conclusions and delineate guidelines for future research.

2 TAT AS A MODEL FOR TEMPORAL ANOMALY DETECTION

The Tunable Activation Threshold (TAT) (Grossman and Paul, 1992) hypothesizes that T-cell activation depends on a threshold that is adjusted dynamically to the integrated history of signals received via the T-cell Receptor (TCR). Every interaction between the TCR and its ligands, the antigenic MHC/peptide complexes presented by the APC, results in intracellular competition between ”excitation” and ”de-excitation” signaling pathways, causing the T-cell to adapt to the stimulus by increasing or decreasing its activation threshold. Therefore, T-cells with different antigen-specificity will have different activation thresholds as they are exposed to different stimuli. Furthermore, Grossman and colleagues also postulated that T-cells that are tuned to be unresponsive by chronic exposure to antigen could inhibit the activation of responsive T-cells in their neighborhood in physical and antigenic spaces. This implies that an immune response will not depend on response of an individual T-cell, but depends on the ensemble of T-cells engaged and on their current activation thresholds, which in turn reflects the T-cell’s individual history.

We have adopted a minimal mathematical model of TAT for T-cells (Carneiro et al., 2005). Briefly, T-cell activation is controlled by two enzymes that respond to antigenic signals delivered by the APC: Kinase (*K*) and Phosphatase (*P*). Antigenic signals 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.

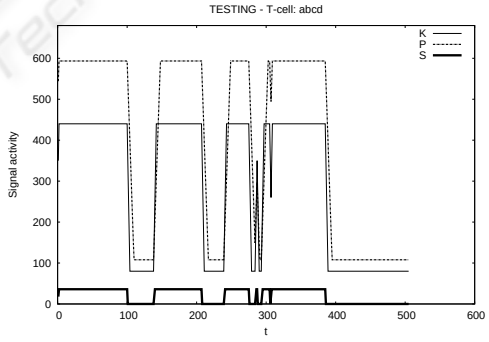

Figure 1: The T-cell receives a variable signal and adjusts the *K* and *P* levels.

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 and those that rarely see their antigen remain sensitive (Carneiro et al., 2005) (illustrated in Figure 1).

3 GENERAL ARCHITECTURE

In this section we present the TAT-based architecture we developed for anomaly detection. We describe a TAT-based AIS and its major metaphorical IS counterparts for anomaly detection.

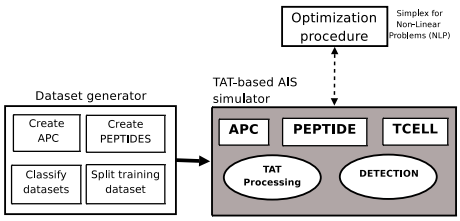

Figure 2: Building blocks of the TAT-based AIS.

Figure 2 depicts the general architecture of TAT-based AIS. The core is a simulation of the dynamics of an artificial T-cell repertoire. Each artificial T-cell, heretofore denoted by *TCELL*, is an object that receives signals from the environment, compares these signals to a *string* representation of its unique specificity, and adjusts its response threshold by tuning the values of its K and P variables. At some given point in time, a "committee" of activated *TCELL*s may raise an alert depending on the values each cell has for these variables.

3.1 Generating Artificial Data-Sets

We have worked with two different data-sets: one is used for *training* and the other is used for *testing*. The training data-set is further split in two parts, respecting the temporal order of the events recollection. The first part is used for *training calibration* in which the T-cell simulator is run, neglecting any *alerts*. The second part is composed by normal and abnormal *APC*s, where the abnormal *APC*s contain peptides not present on the first part of the training data-set. The aim of the training phase is to build and tune the *TCELL* repertoire and adapt the values of K and P to the environment. The second part of the training data-set is used to validate the system initial calibration. This is done by evaluating how well the newly constructed cell repertoire copes with the *APC*s present in the second part of the training data-set.

Each *APC* is a container composed by a set of string *PEPTIDE*s separated by a white space and a *classification tag*. On the experiments described in this paper, the data-sets have been synthetically generated by having the *PEPTIDE*s for each *APC* taken from a group of pre-arranged "string" sets, as follow: a set of normal *PEPTIDE*s (strings) that appear regularly on both training and testing data-sets; a set representing sporadic patterns that appear in training, but are also considered normal; two sets of patterns corresponding to anomalies in training and testing respectively; finally, a set of new patterns for testing that were unseen during training, but are still the result of normal activity.

We have generated (Section 4) artificial data-sets that meet the following conditions: the training data-set has 5000 *APC*s (3750 for calibration and 1250 for validation); the testing data-set has 7500 *APC*s. The *APC*s have a maximum number of 1000 *PEPTIDE*s, generated randomly from the sets described above. The *APC*s with anomalies are different from those generated for the *testing* and for the *training* data-sets.

3.2 *TCELL* Repertoire Dynamics

The AIS contains a variable list of *TCELL*s that are dynamically created and deleted. Each *TCELL* has a unique string, which defines its specificity, and is analogous to the TCR of the natural T-cell. It also stores two variables, K and P, that are adjusted as a function of the input signal S received from each *APC*, as described in next section. A *TCELL* is created and added to the repertoire whenever a *PEPTIDE* in any of the currently queued *APC* does not find a sufficiently similar match in the available repertoire. In this newly created *TCELL* the string is set to be identical to the unmatched *PEPTIDE*. The K and P are initially set to the basal values (K_0 and P_0, respectively) and updated with the stimulus represented by the *PEPTIDE*.

A *TCELL* is removed from the repertoire whenever the K and P dynamics bring them back to these basal values. In practice, this algorithm of creation and removal of *TCELL*s, uses implicitly a potential infinite repertoire of *TCELL*s with K and P in basal values, but we only use the processing and memory resources upon demand, keeping the actual repertoire size contained.

3.3 *TCELL* K and P Dynamics

The TAT model we implemented is a piece-wise linear approximation to the differential equations model described in (Carneiro et al., 2005), and can be described as follows:

1. T-cells are born with basal values of K and P, respectively $K_0 = S_0 * K_{max}$ and $P_0 = S_0 * P_{max}$. S_0 corresponds to the initial value for the signal received by T-cells.
2. The values of K and P are adjusted dynamically as a function of the signal S and tend towards the values $K_0 + S * K_{max}$ and $P_0 + S * P_{max}$, respectively.

3. The input signal S sent by each $PEPTIDE$ to the $TCELL$s in the repertoire is calculated by:

$$S = C \times Affinity(TCELL, PEPTIDE)$$

where C is the number of occurrences of the $PEPTIDE$ in the APC and $Affinity$ is the percentage of equal characters in the same positions, for all the $TCELL$ and $PEPTIDE$ strings.

4. If the values of K and P are lower (higher) than their maximum values, they increase (decrease) linearly with constant derivatives ϕK and ϕP, respectively, until the corresponding plateau values are reached (step 2).
5. The $TCELL$ is transiently activated when K ¿ P; otherwise it is said to be unresponsive.
6. To ensure that a $TCELL$ receiving a constant signal eventually becomes unresponsive (Figure 1) we impose $P_0 + S * P_{max} > K_0 + S * K_{max}$.
7. To ensure that, in any $TCELL$, the condition K ¿ P can be potentially reached at least transiently, we impose that $\phi K > \phi P$.
8. The time duration of each APC is measured in units of $APC_{duration}$ (Δt).

Algorithm 1 shows the corresponding pseudo-code for the updating of $TCELL$ variables.

Algorithm 1 Update $TCELL$ parameters.

if $((S+S_0) * K_{max}) > K$ **then**
  $K \leftarrow MIN((S+S_0) * K_{max}, K+ = \phi K * \Delta t)$
else
  $K \leftarrow MAX((S+S_0) * K_{max}, K- = \phi K * \Delta t)$
end if
if $((S+S_0) * P_{max}) > P$ **then**
  $P \leftarrow MIN((S+S_0) * P_{max}, P+ = \phi P * \Delta t)$
else
  $P \leftarrow MAX((S+S_0) * P_{max}, P- = \phi P * \Delta t)$
end if

3.4 The Immune Response in the AIS

In the AIS, the APCs currently queued are processed sequentially, reflecting the temporal order of events, and are classified according to the activation state of the $TCELL$s that match its $PEPTIDE$s. A $TCELL$ is considered to match a $PEPTIDE$ if their pairwise $Affinity$ is greater than a predefined value α. The classification of the APC is decided based on the *committee* of all $TCELL$s matching its $PEPTIDE$s. We first compute the fraction of activated $TCELL$s per $PEPTIDE$, and count the number of $PEPTIDE$s in which this fraction is greater than a critical value τ. If the number of such 'abnormal' peptides relative to the number of $PEPTIDE$s in the APC is higher than a predefined parameter ψ, then an alert is raised against the APC.

3.5 Adjusting the System

The parameters controlling the natural IS have been slowly refined by millions of years of selection of ancestrals who managed to defend themselves from pathogens, and yet avoided autoimmunity. Similarly, we *set the run-time parameters* of the TAT algorithm by running a non-linear meta-heuristic simplex optimizer (Pedroso, 2007). The mission of this optimizer is to make sure that the TAT-based AIS classifies properly the APCs generated over an appropriate training data-set, tuning automatically the $TCELL$ repertoire to the environment. The optimizer uses only the APCs generated during the training validation to compare the classification made by the AIS algorithm with the classification tag. The optimizer runs the AIS algorithm repeatedly over the bipartite data-set until it finds the parameter regime in which the TAT algorithm tunes the repertoire and is able to raise alerts on the APCs containing artificial anomalies, with a minimal number of false alerts on other APCs.

The introduction of *at least* one artificial anomaly in the training set for parameters validation is absolutely necessary to constrain the tuning dynamics to meaningful parameter regimes. In our experiments we observed that if we would only use normal events in training data-set and require minimization of false alarm rates during the training validation, the simplest solutions returned by the optimizer are parameter regimes in which tuning of K and P is so strong that no $TCELL$ could ever be activated. This allowed us 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. In addition we introduced a few additional heuristic constraints:

$$0 < \frac{K_0}{P_0} < 1 \qquad \phi K > \phi P \qquad 0 < \frac{K_{max}}{P_{max}} < 1$$

The values for K_{max}, ϕK, K_0 and S_0 have also been fixed at 10, 15, 100 and 10 respectively. Also, the values of α, τ and ψ were also optimised.

4 TESTING THE TAT-BASED AIS

Our working hypothesis is that a TAT-based system with the architecture just described could be optimised in such a way that its $TCELL$ repertoire would

be tuned to the individual characteristics of a real environment and yet be able to raise alerts against anomalous activity. In that sense, the main achievement is the development of an implementation of the framework we have previously presented. We have already obtained some results with our implementation, which are described and discussed in the following sections.

4.1 Experimental Protocol

We conducted two sets of five experiments each, with different data-sets. In the first set, each anomaly is extended between two contiguous *APC*s. Our aim is to detect the region (one or both *APC*s) that overlaps with the anomaly. In the second experiment we intend to decide whether TAT could also be used as an *APC classifier*. To be able to determine this, in the second data-set, anomalies appear isolated within a single *APC*. We fixed the maximum number of *APC*s with anomalies in 200 in the training data-set. In the testing data-set we started with 149 anomalies and increase this value in the subsequent experiments. Table 1 describe the optimised parameters sets for each of the data-sets, using the fixed parameters described in subsection 3.5. The upper 5 rows correspond to the "detection" of contiguous *APC*s and the lower 5 rows correspond to the best parameters that correctly classified the *APC*s with anomalies.

4.2 Results

In order to evaluate the characteristics of TAT for detection, let us assume that C is the percentage of rare peptides in the *APC*. In table 2 we consider that the parameter ψ was optimised to the value of 6%, which means that if the ratio of abnormal peptides in the *APC* is above this value, then the *APC* is considered abnormal and an alarm should be raised. For each *APC* we show the concentration of each *PEPTIDE* and the decision made by the AIS. In the bottom we present the classification of each *APC*.

In this example, both *APC*s 2 and 3 have abnormal *PEPTIDE*s and thus should both be *classified* as "abnormal". Nevertheless, since the abnormal *PEPTIDE*s doesn't match any T-cell in the repertoire, new ones are created with the initial values of K and P. According to TAT dynamics (Section 3.3), the signal S (peptide concentration times the affinity) sent by the *APC* should be such that K become higher that P. Thus, the *region* where the anomaly took place comprises the *APC*s 2 and 3 and the AIS raised an alarm in the *APC* 3. The two *APC*s did not have been correctly classified, but the *region* where the anomaly happened

Table 1: TAT optimised parameter set.

Run	P_0	P_{max}	ϕ_P	α	τ	ψ
1	188.4	18.8	2.2	11.1	6.4	88.3
2	127.4	12.7	9.3	38.1	10.1	13.8
3	138.9	13.1	7.9	6.9	2.0	78.3
4	125.1	12.5	9.4	72.9	15.0	54.4
5	144.5	14.4	5.6	57.5	7.3	56.6
1	131.3	13.1	5.8	16.1	8.5	67.9
2	137.1	13.7	3.2	4.32	2.71	77.8
3	127.6	12.7	2.6	80.1	11.8	83.3
4	124.7	12.4	7.3	77.1	10.2	48.7
5	151.4	15.1	4.1	10.4	6.83	55.9

Table 2: Artificial immune detection. ψ = 6%.

PEPTIDE	APC_1		APC_2		APC_3		APC_4	
abcde	84	N	33	N	37	N	107	N
fghij	97	N	53	N	45	N	99	N
klmno	101	N	41	N	36	N	79	N
pqrst	89	N	40	N	39	N	99	N
uvwxz	97	N	53	N	42	N	101	N
PQRST	-	-	43	N	40	Y	-	-
UVWXZ	-	-	36	N	29	Y	-	-
OOOOZ	-	-	43	N	29	Y	-	-
C(%)	0		0		32%		0	
Decision	N		N		**Y**		N	

was correctly detected.

With the resulting optimised parameters sets for the experiences (Table 1), and using the "committee" classification algorithm (Algorithm 1), we obtained the results described in Table 3.

Table 3: Results obtained during the experiments.

	Training			Testing			
Run	*APC*	TP	FP	*APC*	TP qty	FP qty	FP %
1	192	121	15	149	75	14	0.1
2	198	129	20	199	104	66	0.8
3	195	122	17	248	131	34	0.4
4	194	127	18	299	163	83	1.1
5	198	126	23	347	184	31	0.4
1	192	192	18	149	148	21	0.2
2	198	198	12	199	198	46	0.6
3	195	195	159	248	248	563	7.5
4	194	193	92	299	299	417	5.5
5	198	195	27	347	347	35	0.4

Each row represents the best results obtained with the parameters presented on Table 1, for each experiment. The upper 5 rows correspond to the experiments with two contiguously generated anomaly *APC*s and the lower 5 with isolated anomaly *APC*s. In the first experiments, the TAT algorithm could detect at least one of the *APC* carrying the anomaly, which appeared contiguously in the testing data-set. As can be seen, the true positives corresponds to more

than half of the total *APC*s, which means that all the anomalies was detected and, in some cases, both *APC*s of an anomalies raised an alert. We was also succeeded in the last 5 experiments, whose goal was to evaluate the use of TAT for classification. In the testing phase we have a high rate of true positive and a relatively low rate of false alarms. The higher level of false alarms was 7.5% in the third experiment.

5 DISCUSSION

We have described an algorithm for anomaly detection based on the TAT theoretical immunological hypothesis. Our main goal was to present a general architecture of a TAT based AIS and an immune-inspired algorithm for anomaly detection that could deal with temporal events. We presented some preliminary results obtained with artificially generated data-sets that meet some of the characteristics observed on real-world contextual data-sets. We have also started to analyse the appropriateness of using TAT in both a detection and classification context.

Despite the limited diversity of the data-sets used, we believe that the algorithm proposed show that TAT possesses a handful of promising properties when applied to temporal anomaly detection. Firstly, each environment has its own characteristics and therefore the detection system should reflect this individuality, through the automatic adjustment of each cell activation threshold. Secondly, *TCELL* activation is an automatic process based on changes in signal intensity and the current values for the K and P. Each *TCELL* has its sensitivity adjusted to a baseline that is characteristic of the past and current activity. Finally, in TAT, normal activity is manifested by the presence of recurrent signals and abnormal activities correspond to exceptional signals for which the repertoire of *TCELL*s should not be adjusted. This is precisely what is supposed to happen in the detection of anomalies in real-world applications.

In this phase we neglected some essential properties of the natural IS that can make adaptation to evolving normality and anomaly detection more robust and reliable: immune memory and clonal dynamics. Future developments of the TAT-based detector should aim at incorporating these properties. Less because this would make the AIS in line with the natural counterpart, but because variation in clonal size can be a way of adjusting the weight of each *TCELL* specificity in the ”committee”, reflecting not only the history of the signals but also the history of co-occurrences of those signals.

The preliminary results obtained are in line with those described by the authors in (Antunes and Correia, 2008). The results were also promising and the ongoing research give us confidence to deploy a TAT-based algorithm for anomaly detection.

ACKNOWLEDGEMENTS

The authors acknowledge the facilities and research environment gracefully provided by the CRACS (Center for Research in Advanced Computing Systems) research unit, an INESC associate of the Faculty of Science, University of Porto.

REFERENCES

Antunes, M. and Correia, M. (2008). TAT-NIDS: an immune-based anomaly detection architecture for network intrusion detection. IWPACBB08 - Advances in Soft Computing (Springer), pages 6067.

Carneiro, J., Paixo, T., Milutinovic, D., Sousa, J., Leon, K., Gardner, R., and Faro, J. (2005). Immunological self-tolerance: Lessons from mathematical modeling. Journal of Computational and Applied Mathematics, 184(1):77100.

de Castro, L. and Timmis, J. (2002). Artificial Immune Systems: A New Computational Intelligence Approach. Springer.

Forrest, S., Perelson, A., Allen, L., and Cherukuri, R. (1994). Self-nonself discrimination in a computer. Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy, pages 201212.

Greensmith, J., Twycross, J., and Aickelin, U. (2006). Dendritic cells for anomaly detection. In Proc. of the IEEE World Congress on Computational Intelligence, pages 664671.

Grossman, Z. and Paul, W. (1992). Adaptive cellular interactions in the immune system: The tunable activation threshold and the significance of subthreshold responses. Proceedings of the National Academy of Sciences, 89(21):1036510369.

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):413466.

Pedroso, J. (2007). Simple Metaheuristics Using the Simplex Algorithm for Non-linear Programming. LNCS, 4638:217221.

Sompayrac, L. (2008). How the Immune System Works. Blackwell Publishing.