Towards an Immune-Inspired Temporal Anomaly Detection Algorithm Based on Tunable Activation Thresholds

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, are usually based on some form of pattern search on a database of pre-computed "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 and classified or, with the more robust systems, anomalies that are similar to some previously seen event. To cope with these weaknesses there is a new emerging type of anomaly detection system that is "behaviour" based. Although conceptually more appealing, the deployment of behaviour based systems has resulted in an impractical high rate of false alarms.

The vertebrate Immune System is an emergent and appealing metaphor for new ideas on anomaly detection. There are some theories already adopted 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 with training and testing using some synthetic data-sets.

1 Introduction

The vertebrate Immune System (IS) (Sompayrac, 2008) 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. Each immune cell initiates an immune response when it becomes active. This state is dependent on the affinity level the cell has with the pathogen and on its current activation threshold.

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 system 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, grafts and even tumors. 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). The IS is therefore an appealing inspiration to the deployment of Artificial Immune Systems (AIS) for anomaly and intrusion detection (de Castro and Timmis, 2002).

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 Singer, 1996). 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 autoreactive but recognise microorganism structures have low activation thresholds and are thus fully responsive upon infection. We therefore hypothesised that TAT could be successfully applied to temporal anomaly detection, such as network intrusion detection, by adjusting continuously to the temporal sequences of normal network activity, and remaining sensitive to abnormal traffic events. These are the main sources of inspiration and reasoning behind the new bio-inspired anomaly detection algorithm we are currently developing.

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 and 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 The TAT concept as a model for temporal anomaly detection

The Tunable Activation Threshold (TAT) (Grossman and Paul, 1992; Grossman and Singer, 1996) 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 antigens 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.

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.

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 of normal and abnormal APCs, where the abnormal APCs 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 APCs present in the second part of the training data-set.

Each APC is a container composed by a set of string PEPTIDES 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 peptides for each APC taken from a group of pre-arranged “string” sets as follows:

- \( P_N = \{ \text{abcdef, fgij, klmn, pqrs, tuvwz} \} \)
- \( P_O = \{ \text{afkqu, bgkv, aabc} \} \)
- \( P_A = \{ \text{ABCDE, FGHIJ, KLMNO} \} \)
- \( P_B = \{ \text{PQRST, UVWXY, OOOOZ} \} \)
- \( P_U = \{ \text{chmsy, dinsy, ejotz} \} \)

\( P_N \) is the set of normal PEPTIDES (strings) that appear regularly both on training and testing data-sets. We name \( P_D \) the peptide set representing sporadic patterns that appear in training, but are also considered normal. \( P_A \) and \( P_B \) are the sets of patterns corresponding to anomalies in training and testing respectively. Finally, \( P_U \) is the 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 APCs (3750 for calibration and 1250 for validation); the testing data-set has 7500 APCs. The APCs have a maximum number of 1000 PEPTIDES, generated randomly with the patterns described above. The APCs 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 TCELLs 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 \textit{TCELL} the string is set to be identical to the unmatched \textit{PEPTIDE}. The \textit{K} and \textit{P} are initially set to the basal values (\(K_0\) and \(P_0\), respectively) and updated with the stimulus represented by the \textit{PEPTIDE}.

A \textit{TCELL} is removed from the repertoire whenever the \textit{K} and \textit{P} dynamics bring them back to these basal values. In practice, this algorithm of creation and removal of \textit{TCELLs}, uses implicitly a potential infinite repertoire of \textit{TCELLs} with \textit{K} and \textit{P} in basal values, but we only use the processing and memory resources upon demand, keeping the actual repertoire size contained. This optimization saves much processing time and memory resources, allowing for an effective detection with a reasonable small number of \textit{TCELLs} in the simulation repertoires.

### 3.3 \textit{TCELL K} and \textit{P} dynamics

Each \textit{TCELL} receives signals from the \textit{APC} and tunes its threshold accordingly. The \textit{TAT} model we implemented is a piece-wise linear approximation to the differential equation model described in (Carneiro et al., 2005), and can be described as follows:

1. T-cells are born with basal values of \textit{K} and \textit{P}, respectively \(K_0 = S_0 \times K_{max}\) and \(P_0 = S_0 \times P_{max}\).

2. The values of \textit{K} and \textit{P} are adjusted dynamically as a function of the signal \textit{S} and tend towards the asymptotic values \(K_0 + S \times K_{max}\) and \(P_0 + S \times P_{max}\), respectively.

3. The input signal \textit{S} of each \textit{PEPTIDE} is calculated by:

\[
S = \sum_{i=0}^{n} C \times Affinity(\text{TCELL}[i], \text{PEPTIDE})
\]

where \(C\) is the number of occurrences of the \textit{PEPTIDE} in the \textit{APC} and \textit{Affinity} is the percentage of equal characters in the same positions, for all the \textit{TCELL} and \textit{PEPTIDE} strings.

4. If the values of \textit{K} and \textit{P} are lower (higher) than their asymptotic values, they increase (decrease) linearly with constant derivatives \(\phiK\) and \(\phiP\), respectively, until the asymptotic values are reached.

5. The \textit{TCELL} is transiently activated when \(K > P\); otherwise it is said to be unresponsive.

6. To ensure that a \textit{TCELL} receiving a constant signal eventually becomes unresponsive (fig.1a) we impose \(P_0 + S \times P_{max} > K_0 + S \times K_{max}\).

7. To ensure that, in any \textit{TCELL}, the condition \(K > P\) can be potentially reached at least transiently, we impose \(\phiK > \phiP\).

8. The time duration of each \textit{APC} is measured in units of \(\text{APC}_{\text{duration}}(\Delta t)\).

Algorithm 1 shows the corresponding pseudo-code for the updating of \textit{TCELL} variables.

### Algorithm 1 Update \textit{TCELL} parameters.

1. \textbf{if} \((S + S_0) \times K_{max} > K\) \textbf{then}
2. \hspace{1em} \(K \leftarrow \text{MIN}((S + S_0) \times K_{max}, K + = \phiK \times \Delta t)\)
3. \textbf{else}
4. \hspace{1em} \(K \leftarrow \text{MIN}((S + S_0) \times K_{max}, K - = \phiK \times \Delta t)\)
5. \textbf{end if}
6. \textbf{if} \((S + S_0) \times P_{max} > K\) \textbf{then}
7. \hspace{1em} \(K \leftarrow \text{MIN}((S + S_0) \times P_{max}, P + = \phiP \times \Delta t)\)
8. \textbf{else}
9. \hspace{1em} \(P \leftarrow \text{MIN}((S + S_0) \times P_{max}, P - = \phiP \times \Delta t)\)
10. \textbf{end if}

### 3.4 The immune response in the AIS

In the AIS, the \textit{APCs} currently queued are processed sequentially, reflecting the temporal order of events, and are classified according to the activation state of the \textit{TCELLs} that match its \textit{PEPTIDES}. A \textit{TCELL} is considered to match a \textit{PEPTIDE} if their pairwise \textit{Affinity} is greater than a predefined value \(\alpha\). The classification of the \textit{APC} is decided based on the committee of all \textit{TCELLs} matching its \textit{PEPTIDES}. We first compute the fraction of activated \textit{TCELLs} \((C_{\text{active}})\) per \textit{PEPTIDE}, and count the number of \textit{PEPTIDE}s in which this fraction is greater than a critical value \(\tau\). If the number of such 'abnormal' \textit{peptides}, denoted \(C_{\text{abnormal}}\), relative to the number of \textit{PEPTIDE}s in the \textit{APC} is higher than a predefined parameter \(\psi\), then an alert is raised against the \textit{APC}.

### 3.5 Adjusting the system

The parameters controlling the natural IS have been slowly refined by millions of years of selection of ancestors who managed to defend themselves from pathogens, and yet avoided autoimmunity. Similarly, the parameters of the AIS based on \textit{TAT} must also be refined so that the \textit{TCELL} repertoire be automatically tuned to the environment. This section describes the parameters optimization process for setting the run-time parameters of the AIS.

We set the run-time parameters of the \textit{TAT} algorithm by running a non-linear meta-heuristic simplex optimizer (Pedroso, 2007). The mission of this optimizer is to make sure that the \textit{TAT}-based AIS, described in previous sections, classifies properly the \textit{APCs} generated over an appropriate training data-set.
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 APC containing the artificial anomaly 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. If we would only use normal events 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 was what we observed during our preliminary experiments, and was the main reason behind the introduction of anomalies into the training data-set. 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:

$$1 < \frac{P_0}{K_0} < 2 \quad 0 < \frac{\phi K}{\phi P} < 1 \quad 1 < \frac{K_{\text{max}}}{P_{\text{max}}} < 2$$

The values for $K_{\text{max}}$, $\phi K$ and $S_0$ have also been fixed at 10, 18, 10 and 2 respectively. Also, the values of $\alpha$, $\tau$ and $\psi$ 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. Our objective is to explore and develop a TAT-inspired algorithm. In that sense, the main achievement thus far is the development of an implementation of the framework we have presented thus far. We have have already obtained some results with our implementation. These 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 the anomaly comprises always 2 contiguous APCs and we aim to detect the region (one or both APCs) corresponding to the anomaly. In the second set we intend to evaluate if TAT can also be used to classify the APCs. In this context, the APCs with anomalies appear isolated in a non-contiguous way. We fixed the maximum number of APCs 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 APCs and the lower 5 rows correspond to the best parameters that correctly classified the APCs with anomalies.

<table>
<thead>
<tr>
<th>Run</th>
<th>$P_0$</th>
<th>$P_{\text{max}}$</th>
<th>$\phi P$</th>
<th>Aff.</th>
<th>$C_{\text{active}}$</th>
<th>$C_{\text{abn}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>188.4</td>
<td>22.8</td>
<td>15.4</td>
<td>12.5</td>
<td>5.6</td>
<td>88.3</td>
</tr>
<tr>
<td>2</td>
<td>127.4</td>
<td>12.4</td>
<td>18.1</td>
<td>14.4</td>
<td>8.3</td>
<td>77.8</td>
</tr>
<tr>
<td>3</td>
<td>138.9</td>
<td>13.9</td>
<td>15.2</td>
<td>11.2</td>
<td>10.0</td>
<td>54.4</td>
</tr>
<tr>
<td>4</td>
<td>125.1</td>
<td>9.9</td>
<td>13.5</td>
<td>8.1</td>
<td>16.0</td>
<td>54.4</td>
</tr>
<tr>
<td>5</td>
<td>144.5</td>
<td>14.5</td>
<td>9.4</td>
<td>13.5</td>
<td>7.3</td>
<td>56.6</td>
</tr>
</tbody>
</table>

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 $C_{\text{abnormal}}$ 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 APCs 2 and 3 have abnormal PEPTIDEs and thus should both be classified as “abnormal”. Nevertheless, since the abnormal PEPTIDEs don’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 $S$ (the 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 APCs 2 and 3 and the AIS raised an alarm in the APC 3. The two APCs did not have been correctly classified, but the region where
Table 2: Artificial immune detection. $C_{abnormal} = 6\%$

<table>
<thead>
<tr>
<th>Peptide</th>
<th>APC1</th>
<th>APC2</th>
<th>APC3</th>
<th>APC4</th>
</tr>
</thead>
<tbody>
<tr>
<td>abcd</td>
<td>N</td>
<td>33</td>
<td>N</td>
<td>37</td>
</tr>
<tr>
<td>fghj</td>
<td>97</td>
<td>N</td>
<td>53</td>
<td>N</td>
</tr>
<tr>
<td>klmno</td>
<td>101</td>
<td>N</td>
<td>41</td>
<td>N</td>
</tr>
<tr>
<td>pqrst</td>
<td>89</td>
<td>N</td>
<td>40</td>
<td>N</td>
</tr>
<tr>
<td>uvwxz</td>
<td>97</td>
<td>N</td>
<td>53</td>
<td>N</td>
</tr>
<tr>
<td>PQRST</td>
<td>-</td>
<td>-</td>
<td>43</td>
<td>N</td>
</tr>
<tr>
<td>UVWXZ</td>
<td>-</td>
<td>36</td>
<td>29</td>
<td>N</td>
</tr>
<tr>
<td>OOOOZ</td>
<td>-</td>
<td>43</td>
<td>N</td>
<td>29</td>
</tr>
</tbody>
</table>

| C(%)   | 0    | 0    | 32%  | 0    |
| Decision | N    | N    | Y    | N    |

the anomaly happened was correctly detected.

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

Table 3: Results obtained during the experiments.

<table>
<thead>
<tr>
<th>Run</th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>APC</td>
<td>TP</td>
</tr>
<tr>
<td>1</td>
<td>192</td>
<td>121</td>
</tr>
<tr>
<td>2</td>
<td>198</td>
<td>129</td>
</tr>
<tr>
<td>3</td>
<td>195</td>
<td>122</td>
</tr>
<tr>
<td>4</td>
<td>194</td>
<td>127</td>
</tr>
<tr>
<td>5</td>
<td>198</td>
<td>126</td>
</tr>
<tr>
<td>1</td>
<td>192</td>
<td>192</td>
</tr>
<tr>
<td>2</td>
<td>198</td>
<td>198</td>
</tr>
<tr>
<td>3</td>
<td>195</td>
<td>195</td>
</tr>
<tr>
<td>4</td>
<td>194</td>
<td>193</td>
</tr>
<tr>
<td>5</td>
<td>198</td>
<td>195</td>
</tr>
</tbody>
</table>

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 previously collected real-world contextual data-sets, like real network traffic for network intrusion detection. 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 possess 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. In the detection system, the activation threshold of each TCELL is automatically adjusted to the dynamics of real events. 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 that adjusted to a baseline that is characteristic of the past and current activity. Therefore, TAT should more readily cope with the erratic variations in the environment that occur during the operation of real live events. 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 TCELLs should not be adjusted. This is precisely what is supposed to happen in the detection of anomalies in real-world applications.

Some essential properties of the natural IS were neglected in this phase, and their inclusion in future extensions are expected to make adaptation to evolving normality and anomaly detection more robust and reliable. In the natural IS, memory depends also on clonal expansion. Clonal dynamics, either expansion or contraction, is an essential property that we have entirely neglected here. 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 (A and B, 2008).
In this work the goal was to test the use of TAT to detect intrusions in a computer network. The results were also promising and the ongoing research give us confidence to deploy a TAT-based algorithm for anomaly detection.

REFERENCES


