A Network Intrusion Detection System based on the Tunable Activation Threshold theory

Mario J. Antunes and Manuel Eduardo Correia

Technical Report Series: DCC-2007-02

Departamento de Ciência de Computadores
&
Laboratório de Inteligência Artificial e Ciência de Computadores
Faculdade de Ciências da Universidade do Porto
Rua do Campo Alegre, 1021/1055,
4169-007 PORTO,
PORTUGAL
Tel: 220 402 926 Fax: 220 402 950
http://www.dcc.fc.up.pt/Pubs/

A Network Intrusion Detection System based on the Tunable Activation Threshold theory

Mario J. Antunes and Manuel E. Correia
Computer Science Department
Faculty of Science
University of Porto
Rua do Campo Alegre, 823
4150-180 Porto
Portugal
{mantunes;mcc}@dcc.fc.up.pt

Abstract

The main activity of a Network Intrusion Detection System (NIDS) consists in analysing the flow of network packets and identify which ones are part of an ongoing attack or intrusion. Two major problems related with NIDS deployment are the distinction between normal and abnormal activity in the network and the detection of new kind of attacks that have not occurred previously. Several approaches have been applied to solve the problem, with relative success, including machine learning, data mining, statistical and those inspired in the immune system.

In spite of the large body of reseach done on this subject, the literature evidences some problems these approaches have when applied to real world networks. These are mainly due to performance and scalability issues. In this paper we present negative selection and danger theory as two of the major immunological approaches applied so far to the field of intrusion detection. We present what we believe are their major limitations under this context and propose a new NIDS framework based on the Grossman's Tunable Activation Threshold (TAT) theory. This theory is based on the general idea that in the immune system T-cells activation thresholds are adjusted dynamically and this adjustment is based on the recent history of T-cells and APCs interactions.

1 Introduction

An intrusion can be seen as a set of actions that attempt to compromise a secure property. Intrusion detection is the process of monitoring relevant events that occur in a computer-based information system. The main goal of intrusion detection is thus to positively identify all possible occurrences of actual attacks and, at the same time,

to not be distracted by regular events and deceived by the signalling of false attacks [33]. A NIDS has thus to detect unauthorised use, misuse and abuse of computer systems by both system insiders and external intruders.

There are several ways to identify and technologically categorise existing Intrusion Detection System (IDS), such as audit source location and intrusion detection response and detection methods [11]. Considering the source from where an IDS gets its information, these systems can be further classified as Network IDS (NIDS), Host IDS (HIDS) and Hybrids. The intrusion detection response is related with the way the IDS responds to attacks and can be classified as *passive*, *reactive* and *proactive*.

There are also two classes to classify an IDS based on the way it identifies potential intrusions: *anomaly detection* and *misuse detection* [11].

Anomaly (behaviour-based) detection bases its decisions on a profile of normal network or system behavior, denoted by what is called the normal activity profile. The system looks for anomalous activities, which by definition are activities that do not match the previously established profile. An intrusion is thus a deviation from the normal activity profile [4].

The misuse detection (knowledge-based) based systems examines network and system activity, comparing the data collected by the IDS with the contents of a database, looking for known misuses. The database contains the signatures of known attacks in the form of rules. If a match is found, an alert is generated and all the events that do not match any signature are considered not intrusive [4].

Both of these methods of detection have strengths and weaknesses. On one hand, misuse-based systems generally have a very low rate of false positives but cannot identify novel attacks, leading to high false negative rates. On the other hand, anomaly-based systems are able to detect novel attacks but currently produce a large number of both false positives and false negatives [4]. These problems are due to the inability of current anomaly-based techniques to deal adequately with continuous changes in network environments. This is a clear indication for the need to find and apply new paradigms that can better cope with legitimate changes in computer networks and systems usage over time, meaning that any kind of profile for normal behaviour also needs to be dynamic in nature.

The application of biological immune system concepts and algorithms provides the system with the innate capability to distinguish self from non-self, learn new forms of intrusion not previously seen and memorize past events, among other interesting characteristics [10]. These characteristics increase the quality and resilience of these systems by providing them with the ability to react to new and never encountered attacks on networks that change gradually over time.

The immune-based IDS developed so far are generally based on two main immunological theories: Negative Selection (NS) [13] and Danger Theory (DT) [1]. Starting from this well established work we propose a new framework for network intrusion detection based on a different theory proposed by Grossman: TAT [14]. TAT states the activation threshold of Tcells is dynamically adjusted based on the recent history of Tcells-APC interactions. This theory considers a different approach from both NS and DT in what concerns the self-non-self system discrimination.

The selection is made by the T-cell based on the continuous reaction to the

signals received from APC. The activation is based on the tuning of a threshold that reflects the recent history of intracellular interactions between T-cells and APCs. This threshold is not fixed, as stated by NS and the immune response in not based in cell *apoptosis* (death), as proposed by DT. We believe that TAT possesses interesting characteristics that can be applied to IDS and harnessed to define a better metaphor for intrusion detection based on immune systems that is better equipped to cope with the acute problem of effectively detecting intrusions in real-world networks.

In Section 2 we summarize the developments done so far in immune-base IDS using both approaches NS and DT. In Section 3 we describe the fundamentals of the TAT theory and the metaphor that can be thus derived for network intrusion detection. In Section 4 we propose a new work framework based on TAT, describing its main components and processes. In Section 5 we present some conclusions we can derive from this preliminary effort and describe the major ongoing research activities we are currently engaged with to refine these ideas into a fully functional IDS.

2 Artificial Immune Systems applied to IDS

An Artificial Immune Systems (AIS) can be see as an adaptive system inspired by theoretical immunology that can be applied to problem solving [10]. It is widely recognized that network computer security is regarded as one of the most intuitive and popular field of computer science where we can effectively use the biological immune system as a computing metaphor in the form of an AIS.

In [20] the authors present an in-depth description of the state of the art in the development of IDS based on immune biological approaches. The work done so far in this field can be subdivided into three main subareas [20], there are: systems inspired by the immune system that employ conventional algorithms (for example, IBM virus detector from Kephart [17]), those derived from negative selection paradigms, as introduced by Forrest [13] and finally those that take advantage of DT [24]. In this section we give a summary description of the last two approaches, emphasizing their main differences to the TAT theory.

In negative selection [13] the system generates a baseline of *self* patterns based on normal system activity. A large randomly detector set is then generated where each detector is compared to each one of the self patterns. If they match, the detector is destroyed and removed from the initial repertoire. Otherwise, the detector is made available to match the monitored patterns and, if they match with a certain affinity, this should indicate that an abnormal activity has occurred. In her seminal work, Forrest *et al.* [12] managed to take full advantage of some important base characteristics of the immune system, such as diversity, adaptability, anomaly detection and identity by behaviour, among others. In [13] she proposed a first approach to deploy an AIS for network security, where the non-self is characterised as *"undesired network connections"*. In this approach both good and bad connections, as well as the detectors are represented by binary strings. These strings are then subjected to a pattern matching algorithm that is applied to identify self connections. In this first learning phase, the binary strings that are eliminated constitute the negative selection

operation of the AIS being built. On the other hand, if any one of the other surviving patterns matches an antigen and a certain threshold is attained, the corresponding antibody (the pattern matching string) is activated and the presumed intrusion is reported to a human operator that decides if we are truly in the presence of a real incident. If this is the case, the pattern match string is promoted to the memory detector category with the mission to recognize future similar attacks. LISYS [5, 16] was one of the first successful NIDS based on AIS.

In [18] Kim identified three fundamental design goals requirements for network based intrusion detection systems: distribution, self-organisation and lightweight operation. She also concludes a typical AIS framework must include negative selection and clonal selection mechanisms and should take advantage of gene library evolution algorithms. She presents an AIS incorporating the requirements and characteristics listed above, describes the developed architecture and shows some promising results of its application in a real local area network. There are however serious scalability problems associated with the negative selection paradigm when is used in the context of live network traffic [19]. When network traffic increases, the self and non-self space increases dramatically, thus becoming increasingly difficult to find a set of computationally efficient detectors capable of providing adequate coverage of the self and non-self space.

With NS it is no trivial matter to map the entire self and non-self dynamic space. Firstly, they both tend to change over time. Moreover, only some non-self is harmful and one may find some self that can cause damages [18, 1]. More recently, Stibor *et.al* [32] explored the appropriateness of using artificial immune systems based on negative selection for intrusion and anomaly detection problems, specially when compared to other well known statistical anomaly detection methods. In [31] the author identifies some problems related with the use of Hamming shape-spaces applied to anomaly detection in the context of negative-selection based algorithms.

In [3] Aickelin *et al.* presents a survey of the state of the art in intrusion detection systems based on AIS, stressing their weaknesses and defending the need to adopt a new immunological paradigm, the Danger Theory. Matzinger's Danger Theory [24] starts by observing that there must be some kind of discrimination process that goes beyond the classical self-non-self distinction. She bases her argument on evidences from well known natural behaviours. For example, there is no immune reaction to foreign bacteria in the food we eat although they are foreign entities. The human body changes over its lifetime as well but the immune system is still capable of coping with these changes. Other aspects that collide with the traditional viewpoint are the autoimmune diseases which attack the self and successful grafting transplants where there are no attacks against foreign (non-self) tissues. The central idea of the DT is that the immune system does not react to non-self but to danger. The system discriminates "some" self and "some" non-self, which is a starting point to explain why it is possible to cope with "non-self but harmless" and with "self but harmful" system aggressors [2].

The theory states that danger is measured by signals sent out when distressed cells die in some unnatural way. These signals encourage the macrophages to capture antigens in their neighbourhood and establish a *danger zone* around the alarm signal

emitted by the distressed cell. Only those B-cells producing antibodies that match antigens within the danger zone get stimulated and start the clonal expansion process. This theory suggests that the immune system reaction to threats is based on the correlation of various signals, providing a method of linking the threat directly to the attacker. In [1] Aickelin *et al.* transposes the DT to the realms of computer security. Their objective is to specify a computational model based on DT to define, explore and find danger signals. The correlation of danger signals to IDS alerts and these alerts to intrusion scenarios is a subject still far from being completely defined and needs to be better clarified.

In our opinion, this theory has two main drawbacks. First there is the presumption that triggering is based on cell apoptosis. In an IDS implementations this implies that there must have been an intrusion for a correspondent reaction, without a prior prediction that an intrusion is going on. This could be disastrous in a production environment.

Secondly, the meaning and quantification of "danger" can be a hard task of difficult practical applicability to intrusion detection. All computer networks are different, as well as their meaning of what constitute normal and abnormal activities. So is their measure of what constitutes a "danger signal".

3 The Tunable Activation Threshold theory

The biological immune system is a very complex multi-layered structure, composed by a set of cellular components that interact with each other to react against the microorganisms (pathogens), that can cause diseases, such as virus and bacterias. Antigens are substances (usually proteins) identified as foreign by the immune system (the nonself antigens), which stimulates the release of antibodies to destroy pathogens [6]. The immune system is generally divided in two conceptual layers. Firstly, the *innate* immune system, whose behaviour is determined by each person's individual genetic inheritance and responds similarly during each individual entire lifetime. It is composed by a physical barrier (skin), some fluids (e.g. sweat and tears), and once inside the body, by the activity of APCs (for example, the macrophages) that try to destroy the pathogens, fragmenting them into antigenic peptides. Some of these peptides bind to special proteins called Major Histocompatibility Complex (MHC), being presented in the cell surface as a pair "MHC/peptide".

Secondly, the *adaptive* or *specific* immune response, recognizes an antigen as nonself according to prior memory of past intrusions, reacting adaptively to new similar events. In the adaptive system *specificity* refers to the binding process of an antigen (self or nonself) by a cell, in which each cell has a receptor that only recognizes one specific antigen. Furthermore, the molecule surface of an antigen has different antigen peptides that can be bound by different cells. It is therefore possible to have a high number of antigens that can be recognized and destroyed by numerous immune system cells [10].

In the metaphor used in the context of a NIDS the innate immune system corresponds to the baseline knowledge given to the system about know attacks. This can be done by using signatures or rules for well known attacks. This is the approach

normally used by some popular NIDS, like Snort [28].

The adaptive immune system corresponds to the IDS ability to uncover new previously unseen attacks that can occur within the network.

One major mechanism of the immune system is its capability to distinguish self from non-self and thus avoid auto-reactivity. This ability can be partly explained by negative-selection but the problems of coverage, scalability and performance reported by some recent research [19, 32] emphasizes the need for news approaches and paradigms. Another question is related to the usually fixed threshold considered for cell activation and the need for manual intervention in order to confirm the presence of malicious activity.

The Tunable Activation Threshold (TAT) [14, 15] hypothesizes that T-cells have their activation threshold adjusted dynamically by the "kind" and "quality" of signals received from the APCs. This hypothesis proposes that every interaction between T-Cell Receptor (TCR) and its ligands on APC result in an intracellular competition between "excitation" and "de-excitation" signaling pathways, causing the T-cell to increase or decrease its activation threshold [7].

T-cells react differently to the signals they receive from APC (through pairs "MHC/peptide"), *adjusting* its threshold of activation proportionally to the signals received from the APC. Thus, each T-cell has its own responsiveness and tuning updated according to the history of intracellular interactions between T-cell and APC.

The activation threshold increases gradually if the signals received are recurrent and decreases in the absence of signals. T-cells should be *activated* if, in a given period of time, the signals received from the APC are higher than the current threshold. Notice that this can happen if a T-cell does not receive signals from an APC for some time and ends up with a substantially decreased threshold, thus becoming much easier to activate in the presence of higher signals.

The selection process therefore is not pre-programmed, instead, it requires a mechanism of signal-transduction that translates the different external stimulation signals from APC into relatively uniform intracellular signals. Under this scenario threshold signals would allow some of the less excessively autoreactive naive T-cells to survive longer than others, allowing negative selection earlier and late during the maturation process [15]. This behavior contrasts with classical immunology paradigms where all the cells that match the self are naturally discriminated by the survival of those cells that have a higher level of affinity with the non-self (immunocompetent cells).

In TAT the interactions between TCR and "MHC/peptide" complexes induce biochemical changes in the T-cell signalling and activation machinery that alter the sensitivity of T cells to subsequent stimulation [30]. Thus, different T-cells will have different affinity levels to react to the different pairs "MHC/peptides".

The TAT operation model can be described as follows. Each T-cell has two "counters"; One corresponds to the activation threshold (`L`) and the other corresponds to the signal received from the APC in each interaction (`I`). The APC presents a peptide to a T-cell that will adjust its level of activation according to the signal received. Depending on the intensity of the signal, the counter `I` can become higher than `L` and in this case, the T-cell is activated. This mechanism makes the T-cell regulated according to the peptides presented by the APC. The T-cell thus adjusts dynamically

its limits of activations and inaction for a particular antigen. Figure 1 represents the kinetics of intracellular signal intensity and the activation threshold. It illustrates a signal intensity that increases smoothly, adjusting the threshold and a more intense signal that overcomes the threshold limit, which implies the cell activation.

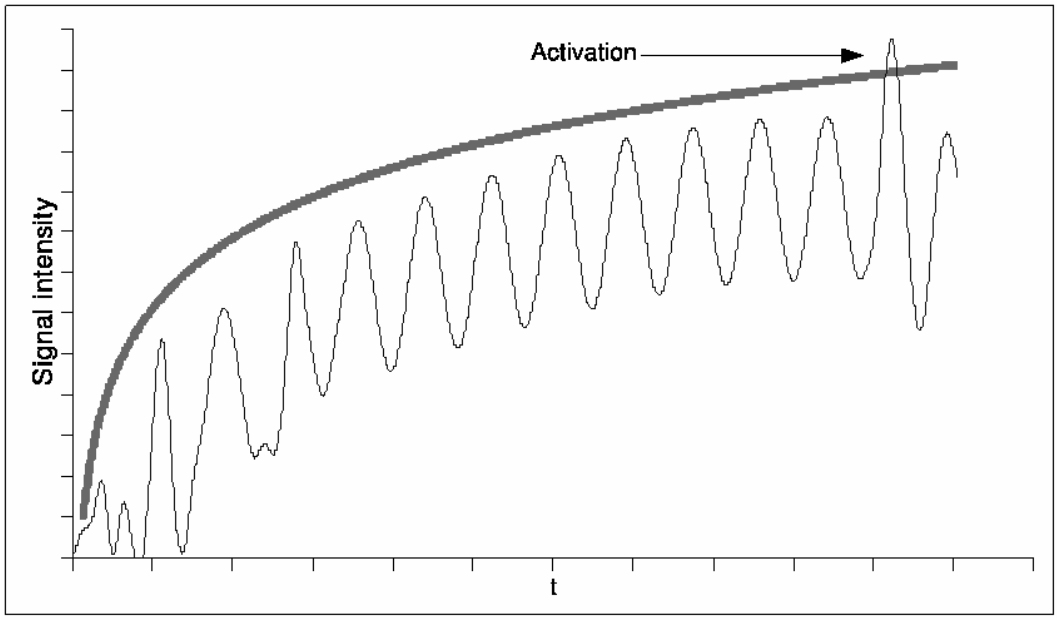

Figure 1: The relation between the signal intensity and the activation threshold of a T-cell.

In TAT the self-non-self discrimination depends heavily on the initial training of the system and the continuous monitoring of the recent history of T-cell-APC interactions. The signal received from the APCs is "self" if all T-cells able to receive the signal have its thresholds (variable `L`) adapted and the signal is below the threshold. On the other hand, if the signal received from the APC is above the threshold of all the T-cells trained to receive it, then the signal is considered by TAT as "non-self".

The system key phase is thus the training phase. The baseline of the system is the normal (self) behavior, and this knowledge is used to produce the TCR. This naive cells are born in the thymus according to each individual genetic information. Its activation threshold is high and during the maturation phase it decreases naturally and spontaneously at a defined rate. In the T-cells that receive recurrent signals (self patterns), the decreasing of the threshold is opposed by a natural tendency for the signal to increase it. In this case, the activation threshold will be always above the input signal received by the APC. On the other hand, in T-cells that do not receive enough signals, the threshold will also decrease, but there is no signal in opposite direction. Thus, in some moment, they will be activated, turning these cells too reactive to detect non-self patterns [14]. This dynamic operation makes the automatic

adjustment of the activation threshold in T-cells dependent on two main factors: their initial training and the monitorization of the signals received by the APC.

The application of TAT to intrusion detection thus has the following points of interest:

- the automatic adjustment of T-cells activation threshold based on the system activity reflects more accurately what really happens in a network. The network traffic is different in all the networks and it is necessary to have detectors compatible with this reality.
- the activation is an automatic process based on the kinetics between the signal intensity and threshold and thus the manual intervention to confirm the attack may no longer be required.
- the dynamic threshold seem to be more realistic than the model defended by classical immunology, as it reflects the real operation of the system. The recognition of a new unseen intruder (an attack) depends on the "strength" of the signal received from the APCs (traffic filter).
- the normal operation of the system should fine tune the threshold of some T-cells, converting them into "memory" cells. For example, when an attack takes place in the network, the T-cells that receive such signals will automatically adjust its threshold to a value that will allow for their activation making that cell reactive to this same attack or some of its variants.
- the gradual threshold adjustment over time tends to minimize (or even eliminate) the false negative events because T-cells will only activate when the signal is above the threshold.
- the activation is triggered when the bind match a threshold adjusted dynamically over time. This should reflect the dynamic history of the system, instead of a pre-defined state supposed to reflect the natural evolution of those individuals. Computer networks are not all equal and each one has its own dynamics for normal activity. Each network should thus adjust dynamically its threshold of reaction according to its own activity profile.
- recurrent signals are usually related to normal activity. This is precisely what usually happens in a network. Abnormal activities are exceptional signals that should adjust the threshold to a level capable of activating the T-cell.

The metaphor of TAT applied to intrusion detection is summarized in the figure 2.

At this phase of our research, there are some questions that need to be better clarified:

- it is not clear what should be the rates in which the threshold (`L`) and signal input variable (`I`) should vary, to reflect a real-world network system. If the signal is recurrent, then `L` should increase more than `I`.

Immune system (TAT)	network environment
Thymus	detector set generator
Tcell	detector
TCR	detector string
activation threshold	Variable L of the detector
signal	Variable I of the detector
Tcell activation	alarm trigered if I>L
APC	traffic filter
Peptide	substrings matched by each APC
peptide concentration	ocurrences of each string filtered by APC
tuning threshold	automatic adjust to variable L of the detector
Self	Normal activities
Non Self	Abnormal activities

Figure 2: The metaphor of TAT and the intrusion detection.

- the strings that each APC will match in the network traffic must also be very well defined.

In our opinion, the self-non-self distinction proposed by TAT has some very interesting and metaphoric insights that can be easily mapped and applied to intrusion and anomaly detection, when compared to NS and DT. Firstly, the T-cell selection is directly related to its interaction with the environment, avoiding any premature distinction in the *thymus*, as happens with NS. Moreover, the significant differences between both approaches are based on the fact that activation is based on a tunable threshold instead of a fixed one. We may thus have different individual T-cells with different levels of reaction.

Comparing to DT, TAT is not based on cellular apoptosis. So, the system is not expecting to be infected for latter reaction. In TAT, the system detects the intruder if the signal received is above the T-cell threshold. Moreover, there is no need to "classify" each signal since what matters is its intensity and its relation to its recent occurrences level. A precise meaning for what is a "danger signal" and the need for its correlation is not necessary.

4 The proposed framework

In this section we present a framework for an IDS based on the TAT model summarily described in Section 3. Figure 3 illustrates the general proposed architecture,

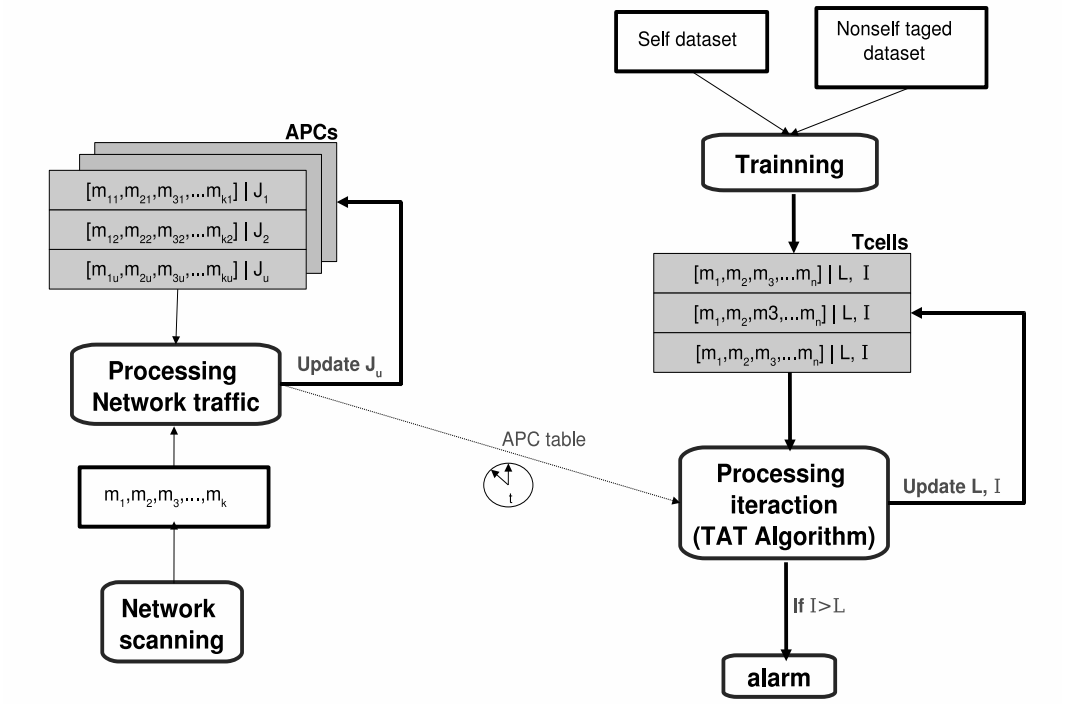

Figure 3: General architecture of the NIDS using TAT

emphasizing its main processes and data flows.

Our system is composed by two main structures that interact periodically: Tcells, corresponding to detectors and APCs corresponding to network filters that extract patterns from network flows. A network communication flow identifies all the packets exchanged between two applications in the network and can be identified by the following attributes: source and target IP address, source and target ports (Transport layer), protocol type (Internet layer), service type (TCP header) and input router interface [29]. We can identify three main operational phases: (1) the initial training of Tcells, (2) network traffic processing and (3) T-cell and APC interactions based on the TAT algorithm. Each detector is identified by a string of `m` attributes and has two variables: the activation threshold (variable `L`) and the intensity of the signal received by each APC interaction (variable `I`). The APCs are identified by a vector of sub-strings with `k` attributes (with $k < m$), corresponding to the traffic being filtered from the network, and a variable `J` corresponding to the number of occurrences for each string.

The system operates as follow:

1. it starts by creating two sets: detectors represented by Tcells and traffic filters represented by APCs.

2. variables `L` are initialized with a very large pre-defined value for all Tcells.

3. variable `I` is reset to zero for each interaction T-cell-APC.

4. APCs collect network traffic in real time and store the occurrences of each of its filters. The number of occurrences is stored in variable `J`.

5. each entry in the APC table is periodically presented to all the TCR and, in case of a match, the variable `I` is updated with the value `J`. The variable `L` is also increased by a value that should reflect the intensity of the signal (`I`). This means that the signals are recurrent, corresponding to self activity, and both variables are incremented in such a way that `L` always becomes greater than `I`.

6. on the other hand, if the detector finds a rare or a too strong signal, then both `I` and `L` will decrease, but `L` is made to decreases faster, causing `I` to be higher than `L` at some time in the future, causing the detector to become activated.

The activity of the APCs can be described as follow:

1. each network filter (APC) extracts several patterns from the network and counts occurrences in a defined period of time.

2. These occurrences are stored in a APC table, associated to each string (variable `J`)

3. In the case that no match exists to a particular string ($J = 0$), then that string is removed from the list.

This description emphasizes the general characteristics described in the previous section, being possible to identify the metaphor proposed in Figure 2 (Section 3).

Figure 4 details the main processes involved. The training is composed by two distinct separate phases. Firstly, the system is trained with a *self* data composed exclusively of normal traffic from the network we want to protect (*Self-Tagged*). Secondly, we train the system with a data-set composed both by normal and known attacks (*Non-self Tagged*). One possible approach is to "synthetically" generate these attacks from the detection rules of the Snort IDS [28]. This second training phase matures T-cells by causing an adaptation of its thresholds and turning them more sensible to these attacks.

This procedure will tune some T-cells to be highly adjusted to react to the already known attacks. These T-cells act as "memory cells" since their existence records previous known attacks. The activation threshold for these cells should also remain in a low value (near zero) in order to make them react quicker in any re-occurrence of these attacks.

The monitorization process can occur in several points of the network. For example, it could consist of processes running on several PCs in a local area network. Each process has a set of generated APCs that will filter the network traffic according to

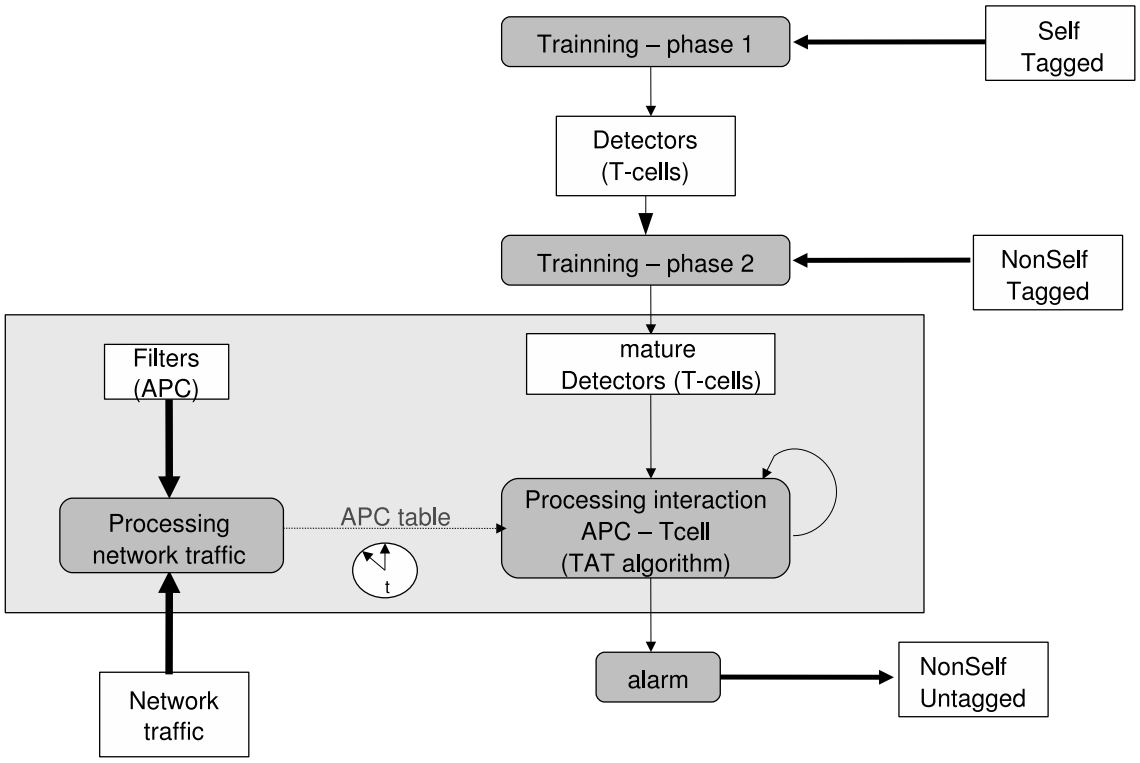

Figure 4: General architecture for the NIDS.

specific rules. These rules will be used to create a vector of sub-strings that should filter in real time the traffic collected from the network. The APC stores, for each sub-string, the number of occurrences thus found.

For network traffic capture ("*sniffing*") we intend to use the netmate [25] tool. This is a flexible and extensible measurement tool written in Java. It has several modules for packet accounting, delay/loss measurement, packet capturing and netflow statistics. We have done several experiments collecting netflows from a medium size network that confirmed the robustness and appropriateness of this tool, mainly its ability to classify network flows and to be easily extended with more specific netflow characteristics, such as the total packets sent in each direction, amount of bytes transmitted and duration among others. During capture, besides the net flow statistics, it is possible to collect several different kind of data from network traffic such as bandwidth measures, packet delay and the payload for each packet.

For our framework, in order to carry on the two training phases, we intend to collect traffic from normal activity of the network, as well as the traffic generated during induced synthetic network attacks. The data thus collected will allow us to produce two very different data-sets for further classification.

To obtain the first preliminary results we intend to use the DARPA/MIT Lincoln Laboratory off-line intrusion detection evaluation data set [23]. This off-line data set contains around 500000 network connections, between normal and 17 labeled attacks.

The amount of characteristics (features) collected by netmate, as well as by some other network monitoring tools, is very large and, in some ways, redundant. From

another perspective, not all the features are really important for intrusion detection [21]. It is possible to reduce the feature set without loosing accuracy. This is largely substantiated by the large body of research done on the use of data mining techniques[22] and machine learning [27, 26, 8, 9] approaches.

A much reduced feature set can generally be obtained from the training data-set, such that: (1) it maintains the accuracy for the classification process and (2) it becomes computationally possible, in real time, to process the data collected by netmate. Also, in terms of representation, the reduced dimension of the feature set can bring good benefits in the representation of the normal data-set in the shape-space (Hamming or real-value), as well as in the measuring of the distance between T-cells and the events observed.

We intend to explore mechanisms of feature set reduction for a better shape-space coverage, starting by some accepted research in the field [9, 27]. The results described in [30], explains the modulation of TAT in an Hamming space-space. We intend to study the application of this model to our research and explore the appropriateness of TAT implemented with Hamming shape-spaces as a network intrusion detection system. We also intend to measure the accuracy obtained with different feature sets and what can be its impact in the overall system performance.

The use of TAT in a NIDS context has the potential of bringing new insights to the AIS research. A working system will certainly help to justify some immunological behaviors that are not yet well understood.

5 Conclusions and future work

We have proposed a novel NIDS framework based on the TAT theory and presented a framework for its application in the context of network intrusion detection. We have also described NS and DT as the two main immunological models applied to AIS-based IDS so far, summarized its behaviour and emphasized their major drawbacks when applied to real large scale network intrusion detection.

The TAT theory sustains that the activation of a T-cell (network detector) is based in a threshold that is fine tuned according to the recent history of Tcell-APC interactions. This general idea gives interesting insights to the application of TAT to practical NIDS.

In [20] it is argued that the research in AIS-based IDS, and its experimental results so far, have shown that these systems are only able to work on relatively simple, small problems, in very selected environments. The authors also emphasize the need to explore new immunological mechanisms that have not been previously studied and applied for intrusion detection. The theoretical study of TAT, the practical results obtained with TAT in other contexts [7] gives us good confidence that this theory can be applied with success in the deployment of effective NIDS.

Our ongoing research will continue with the development and implementation of this architecture in order to obtain comprehensive results about the use of TAT in the context of NIDS. The next steps will be to define a methodology to cover in a more optimal way the events shape-space (Hamming or Real-value), to define and test

functions for adapting the thresholds and to fine tune the training phase for obtaining better insights about the applicability and real value of TAT for and efficient network intrusion detection.

References

[1] U. Aickelin, P. Bentley, S. Cayzer, J. Kim, and J. McLeod. Danger theory: The link between ais and ids? *Proc. of the Second Internation Conference on Artificial Immune Systems(ICARIS-03)*, pages 147–155, 2003.

[2] 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)*, pages 141–148, 2002.

[3] U. Aickelin, J. Greensmith, and J. Twycross. Immune system approaches to intrusion detection-a review. *Proc. of the Second Internation Conference on Artificial Immune Systems(ICARIS-03)*, pages 316–329, 2004.

[4] Y. Bai and H. Kobayashi. Intrusion detection systems: technology and development. *Advanced Information Networking and Applications, 2003. AINA 2003. 17thInternational Conference on*, pages 710–715, 2003.

[5] J. Balthrop, F. Esponda, S. Forrest, and M. Glickman. Coverage and generalization in an artificial immune system. *GECCO-2002: Proceedings of the Genetic and Evolutionary Computation Conference*, pages 3–10, 2002.

[6] G.R. Burmester and A. Pezzuto. *Color Atlas of Immunology*. Thieme Medical Publishers, 2003.

[7] 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*, 184(1):77–100, 2005.

[8] S. Chebrolu, A. Abraham, and J. Thomas. Feature deduction and ensemble design of intrusion detection systems. *Computers and Security, Elsevier Science*, 2005.

[9] R. Cilibrasi and PMB Vitani. Clustering by compression. *Information Theory, IEEE Transactions on*, 51(4):1523–1545, 2005.

[10] L.N. de Castro and J. Timmis. *Artificial Immune Systems: A New Computational Intelligence Approach*. Springer, 2002.

[11] H. Debar, M. Dacier, and A. Wespi. Towards a taxonomy of intrusion-detection systems. *Computer Networks*, 31(8):805–822, 1999.

[12] S. Forrest and S. Hofmeyr. Engineering an immune system. *Graft*, 4(5):5–9, 2001.

[13] S. Forrest, A.S. Perelson, L. Allen, and R. Cherukuri. Self-nonself discrimination in a computer. *Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy*, pages 201–212, 1994.

[14] Z. Grossman and WE 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*, 89(21):10365–10369, 1992.

[15] Z. Grossman and A. Singer. Tuning of activation thresholds explains flexibility in the selection and development of t cells in the thymus. 1996.

[16] S.A. Hofmeyr and S. Forrest. Immunity by design: An artificial immune system. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, pages 1289–1296, 1999.

[17] J.O. Kephart. A biologically inspired immune system for computers. *Artificial Life IV*, 1994.

[18] J. Kim. *Integrating Artificial Immune Algorithms for Intrusion Detection*. PhD thesis, University of London, 2002.

[19] J. Kim and P. Bentley. An evaluation of negative selection in an artificial immune system for networkintrusion detection. *Genetic and Evolutionary Computation Conference 2001*, pages 1330–1337.

[20] J. Kim, P. Bentley, U. Aickelin, J. Greensmith, G. Tedesco, and J. Twycross. Immune system approaches to intrusion detection - a review. *Natural computing*, 2007.

[21] W. Lee and S.J. Stolfo. A framework for constructing features and models for intrusion detection systems. *ACM Transactions on Information and System Security (TISSEC)*, 3(4):227–261, 2000.

[22] W. Lee, S.J. Stolfo, and K.W. Mok. Adaptive intrusion detection: A data mining approach. *Artificial Intelligence Review*, 14(6):533–567, 2000.

[23] M.V. Mahoney and P.K. Chan. An analysis of the 1999 darpa/lincoln laboratory evaluation data for network anomaly detection. *RAID 2003*, pages 220–237.

[24] P. Matzinger. The danger model: A renewed sense of self. *Science's STKE*, 296(5566):301–305.

[25] Netmate Meter. http://sourceforge.net/projects/netmate-meter/. as of April 2007.

[26] S. Moyle and J. Heasman. Machine learning to detect intrusion strategies. *LECTURE NOTES IN COMPUTER SCIENCE*, pages 371–378, 2003.

[27] S. Mukkamala and A.H. Sung. Feature selection for intrusion detection using neural networks and support vector machines. *Journal of the Transportation Research Board of the National Academics, Transportation Research Record*, 1822:33–39, 2003.

[28] Marc Norton and Daniel Roelker. Snort 2.0: High performance multi-rule inspection engine. 2003.

[29] J. Quittek. Rfc 3917-requirements for ip flow information export (ipfix) http://www.ietf. Technical report, org/rfc/rfc3917. txt.

[30] A. Scherer, A. Noest, and R.J. de Boer. Activation-threshold tuning in an affinity model for the T-cell repertoire. *Proceedings: Biological Sciences*, 271(1539):609–616, 2004.

[31] T. Stibor, P. Mohr, J. Timmis, and C. Eckert. Is negative selection appropriate for anomaly detection? *Proceedings of the 2005 conference on Genetic and evolutionary computation*, pages 321–328, 2005.

[32] 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*, 2, 2005.

[33] HS Venter and JHP Eloff. A taxonomy for information security technologies. *Computers and Security*, 22(4):299–307, 2003.