Mário J. Antunes and Manuel E. Correia

Towards a new Immunity-Inspired Intrusion Detection Framework.

Technical Report Series: DCC-2006-4

Departamento de 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 Web: www.dcc.fc.up.pt

Towards a new Immunity-Inspired Intrusion Detection Framework

Mário J. Antunes
Computer Science Department
Faculty of Science
University of Porto
Porto, Portugal
mantunes@dcc.fc.up.pt

Manuel E. Correia
Computer Science Department
Faculty of Science
University of Porto
Porto, Portugal
mcc@dcc.fc.up.pt

Abstract

In this document we introduce a novel framework for behaviour based Network Intrusion Detection Systems (NIDS). Its main goal is the application of theoretical immunological concepts to provide adaptability to the normality of the network behaviour, based on memory and learning from previous attacks. We present some important principles and concepts relevant to the description and categorization of Intrusion Detection Systems (IDS), and then describe the main benefits that can be obtained from an Artificial Immune System (AIS) approach for IDS. We conclude by proposing a novel extension to the Common Intrusion Detection Framework (CIDF) capable of accommodating our initial goals.

1. Introduction

The Network Intrusion Detection System's (NIDS) role consists on analyzing the packets in transit and identifying which ones is part of an attack. There are two major complementary problems related with NIDS deployment, which have been the target of intense research during the last decade: the definition of anomaly, based on the distinction between normal and abnormal traffic; and the detection of new unknown attacks before they can cause serious damage.

To address these issues, several approaches have been published and discussed, but none seems to provide a completely satisfactory answer yet. In this document we propose a new approach to solve these problems based on the implementation of biologically inspired concepts and algorithms, such as the ones related to the human immune system [8][27]. With this approach, researchers aim to develop computer systems that can take advantage of concepts, ideas and algorithms based on the theoretical biological models of the human immune system (AIS - *Artificial Immune Systems* [9][11]), and apply it to intrusion and anomaly detection on computer networks.

Starting with previous and well accepted work on Intrusion Detection Systems (IDS), we propose an extension to the Common Intrusion Detection Framework (CIDF) [12], enhancing its capabilities to work with an adaptable signature of normal traffic. In our development the definition of normal traffic is still open to various approaches, such as statistical, data mining and expert systems, among others approaches ([29][32][37][38]).

In sections 2 and 3 we introduce the fundamentals of IDS and biological immune system, respectively. In section 4 we present relevant work related to artificial immune system models that have been applied to intrusion detection systems. The proposed system, I_2NF (***I****mmunity-****I****nspired* ***N****etwork Intrusion Detection System* ***F****ramework*), is described and analyzed in section 5, where we also present the CIDF and a new proposal that gives adaptability to the IDS.

Finally, in section 6, we present some conclusions reflecting the research we have done so far and discuss directions for future development.

2. Intrusion Detection Systems

In the last decades the technological, productive and economical efficiency of the organizations has increased in an unprecedented scale. This has been fuelled by the massive deployment of information and communication technologies (ICT), mainly by distributing information and computation power on often crucial business processes. This has inadvertently promoted easy and often insecure access to critical information, which can be used to cause damages by criminal activities or incautious use. One of the main current concerns is thus to monitor and manage the flow of information in computer networks, in order to mitigate the risks associated with what is deemed an *intrusion*. Clearly there is a need for

detection controls capable of discovering intrusions, generally called IDS, which will operate in a very complex medium where millions of benign events hide similar but unwanted and sometimes dangerous activities.

To better contextualize the discussion of this topic, we now proceed with the presentation of some of the basics of information security and then continue with a brief description of what constitutes an IDS.

2.1 Fundamentals of information security

A secure computer system guarantees (through adequate controls) one or more of the following fundamental properties [28][33][39][40]:

- *Confidentiality*: the information is only accessed by who has official permission.
- *Integrity*: the information is not modified without official permission.
- *Availability*: the information is accessible to legitimate entities when required.

Specific environments may require more secure properties, like authority and non-repudiation, but this discussion is not relevant here.

To protect themselves from potentially damages resulting from illicit activities, organizations can defend themselves by putting into practice security policies. These policies provide the rules to secure the information stored and processed by the organization's ICT. In short, the *security policy* [7] defines the main rules for the access and manipulation of information by users, systems administrators and organization employees.

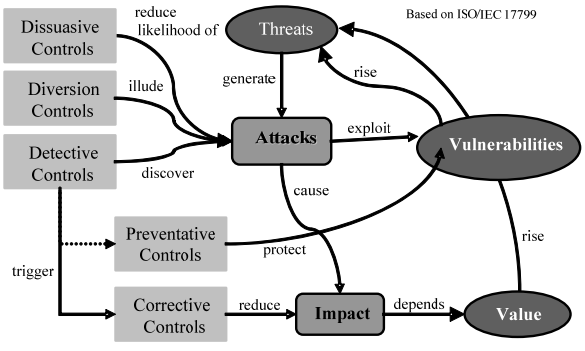

Figure 1 - Security policy model based on ISO/IEC 17799.

Figure 1 illustrates a possible model for the security policy development process [33]. In this model a *vulnerability* [31] is a flaw (know or unknown) in the information system that can be exploited. A *threat* is the potentiality of a deliberate and unauthorized attempt that can cause damage. An *attack* is a specific execution of a plan to carry out a threat that exploits one or more vulnerabilities. The *impact* is the repercussion of a successful attack. The impact allows for the quantification and assessment of the damage inflicted to the information system by the attack. To minimize the risk, the security policy provides mechanisms to improve security, generally falling within one of the following categories:

- *Dissuasive actions*: to minimize the probability of an occurrence of a successful attack.
- *Prevention actions*: to protect previously identified vulnerabilities.
- *Corrective actions*: to minimize the impact of an attack.
- *Detection actions*: to detect and identify effective attacks –IDSs naturally fall within this class.
- *Diversion actions*: to deceive an intruder with a forged target, thus hiding the true vulnerability.

All these mechanisms can be implemented by several technologies, such as firewalls, anti-virus, honeypots and IDS.

2.2 Intrusion Detection and IDS

By *intrusion* [34] we mean a set of actions that attempt to compromise a secure property. *Intrusion detection* [43] 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 occurrences of true attacks and, at the same time to not be mistaken about regular events, signalling false attacks. The IDS generally includes technology to automate this monitoring and analysis process. Bai [4] provides a summary of the different tasks performed by an IDS, namely:

- Monitoring and analyzing user and system activities.
- A *recognition activity model*, mapping known attacks and alerts.
- Statistic analysis model for abnormal behaviour detection.
- Evaluating system and data files integrity.

2.3 Metrics to evaluate the quality of an IDS

There are several metrics available to evaluate IDSs' operation [4] [17], namely:

- *Accuracy:* an inaccuracy occurs when an IDS flags as intrusive a legitimate action (false positive). The number of false positives (false alarms) defines the accuracy of an IDS.
- *Completeness:* IDS should not fail to detect an intrusion. However, it is difficult to fulfil this requirement because it is almost impossible to have a global knowledge about all possible types of attacks. A false negative occurs when the IDS fails to detect a real attack.
- *Effectiveness*: system's ability to detect intrusions. The number of false positives and false negatives defines the IDS efficiency.
- *Performance*: refers to the ability of an IDS to perform intrusion detections in real-time.
- *Adaptability*: refers to the ability of an IDS to recognize slight variations of known attacks.
- *Extensibility*: means that IDS can be easily customized.
- *Fault tolerance*: means that IDS itself is immune to attacks, particularly *Denial-of-Service* (DoS).
- *Timeliness*: refers to the performance of an IDS, but also the temporal period required to propagate its events analysis for further reaction.

2.4 Taxonomy for IDS

There are several ways to identify and technologically categorize different types of IDS. Here we present a taxonomy that uses three main characteristics: audit source location, intrusion detection response and detection method [17].

Concerning the audit source location, there are several different places from where an IDS can get its information, namely:

- *Network packets*: the data collection is based on raw network packets. This IDS class is generally known as Network Intrusion Detection System (NIDS). The main advantages of a NIDS include: positive detection of unsuccessful attacks, real time attack detection and the independence from the host operating system (OS). However, NIDS have several drawbacks. First, on segmented networks it may be necessary to put a NIDS in each one of the individual segments. Secondly, a NIDS may have problems in dealing with fragmented packets. NIDS are also unsuitable for encrypted data. Finally, NIDS may have problems monitoring high speeds networks with a high volume of network traffic.

- *Host* and *application log files*: IDSs that use this data source are named Host Intrusion Detection System (HIDS) and they use as input the information received from the OS or applications log files. The main properties of HIDS are: no additional hardware needed; well suited to detect insider attacks; detection of attacks that elude NIDS; well adapted to encrypted data and switched networks. However, HIDS have also several drawbacks: host OS dependence; harder to manage (information must be configured and managed for each host); can be disabled by DoS attacks; and the high data volume generated may require additional storage.

- *Hybrid*: this kind of IDS includes characteristics of both a NIDS and a HIDS.

The intrusion detection response is related with the way IDSs respond to attacks. The taxonomy divides them into three classes:

- *Passive response*: it is executed after the attack has been detected. For example, by notifying the security officer or by sending an alert to a console or an email.
- *Reactive response*: it is also executed after the attack has been detected, but the goal is to mitigate the attack effects.
- *Proactive response*: the response interferes and tries to actively stop an attack from accomplishment.

Finally, the detection method describes how an IDS detects intrusions. There are basically two ways:

- *Behaviour-based*, *anomaly detection* or *detection by behaviour*: is a technique based on the assumption that all intrusive activities are necessarily anomalous. Thus, it first builds a model of the normal behaviour of the system denoted by "normal activity profile". It then looks for anomalous activities, which do not match the previously established profile. An intrusion is flagged by observing a deviation

from the normal activity profile. This anomaly detection method is effective in detecting unknown attacks. However, this method generates a huge number of both false positives and false negatives. There are several approaches for anomaly detection [4], namely statistical analysis, predictive pattern generation, neural networks and genetic algorithms.

- *Knowledge-based*, *signature-based* or *misuse detection*: is based on the description of known attacks by a signature or a pattern, generally referred as a rule. These rules are stored in a database. With these signatures the system tries to find out or identify all intrusive activities. The data collected by the IDS are compared with the database contents and, if a match is found, an alert is generated. All events that do not match any signature are considered not intrusive. With this method, variations of the same attack can not be theoretically detected. The signature-based technique tries to recognize known bad behaviour and has the potential for very low number of false positives. Drawbacks include the difficulty of identifying unknown attacks – high false negatives – and the need for a constant upgrade and demanding maintenance cycle. There are several approaches for misuse detection [4], namely, expert system, state-transition analysis and pattern matching.

2.5 IDSs Challenges

Ideally, a successful IDS must detect all intrusive activities and should not generate false alarms. Unfortunately, this goal is far from being achieved in its entirety. There are several challenges affronting the generalized successful deployment of IDS. These comprise improvements on efficiency, configuration and operation. We identify the following main emergent challenges on IDS deployments:

- *New approaches for intrusion detection*: currently there are good commercial solutions for both signature-based and behaviour-based IDS. However their potential performance for detecting unknown attacks is crippled by a high false alarm rate. To address this issue, a new approach to define *normality*, with the capacity to adapt to the traffic changes and to deal with small variations in the interpretation of events, is required. For example, there is a need to deal with unauthorized activities outside the normal profile previously defined.

- *IDS benchmarking*: presently there is no recognized standard for IDS evaluation metrics. It is thus very difficult to compare the potential characteristics of all IDS. With metric normalization and a common data set, it would be much simpler and more reliable to examine the main IDS requirements and then assess the relative merit of each solution.

- *New detection methods*: some researchers are interested in using data mining, data fusion, and immunological approaches to build detection models for IDS. *Data mining* [4] has a great potential in minimizing the problem by dealing with large amounts of data and can be used on both anomaly and misuse detection. *Data fusion* [4][6] is a method used to combine data from multiple and diverse sensors and sources in order to make inferences about intrusive activities. The *immunological approach* is based on biological analogies (section 3) [9]. This provides the IDS with the ability to react to new and never encountered attacks by taking advantage of the learning and memorization of past events. In section 5 we describe a framework for intrusion detection based on this approach, which has the great potential to provide an IDS with the capability to self adjust in order to detect and/or react to unknown attacks.

3. Biological Immune System

The biological immune system [8][27] protects and defends the human body from intrusions of microorganisms (*pathogens*) that can cause diseases, such as virus and bacteria. *Antigens* are substances (usually proteins) identified as foreign by the immune system. They stimulate the release of antibodies to destroy the pathogens. This section introduces some basic concepts present in the biological description of the algorithmic immune system activity and the inspiration beyond its use as a metaphor in computer applications.

3.1. Basic concepts

The Human immune system is a very complex multi layered structure (Figure 2) [9]. It is composed by a set

of cellular components that interact with each other to react against an intruder.

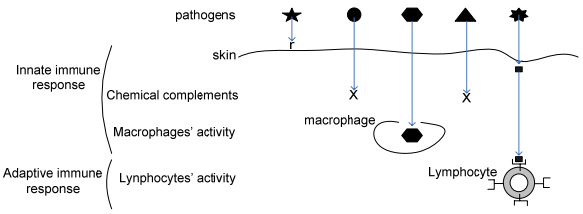

Figure 2 - Multi layered structure of immune system.

After the first line of defence (skin), there is a chemical barrier, where several fluids such as sweat and tears, interact with the pathogens to try to destroy them. Inside the body there exists a third line of defence to deal with pathogens. It uses a specific kind of cell called a *macrophage*. These cells move all over the body seeking for foreign microorganisms and trying to destroy them. These three lines of defence are usually called *innate*, because they act naturally, as a result of each person's individual genetic information. They also act in the same way for all situations during their entire lifetime. On the other hand, the *adaptive* or *specific immune responses* recognize an antigen according to prior memory of past intrusions, reacting adaptively to new or similar events.

In the adaptive system, the specificity refers to the binding process of an antigen 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 a large number of immune system cells.

The leukocytes (white blood cells) are the main cellular components of the immune system, and its taxonomy is illustrated in Figure 3.

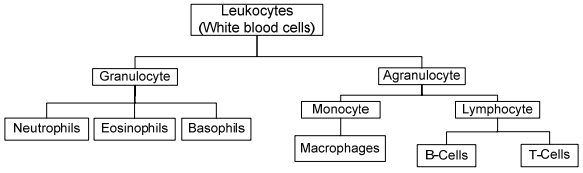

Figure 3 - Taxonomy of Leukocytes.

The leukocytes can contain visible granules (granulocytes) like neutrophils, eosinophils and basophils, or be nongranular (agranulocytes), like the macrophages and the lymphocytes. The macrophages are the most important cells of the innate response, having the responsibility of destroying the foreign cells. Nevertheless with different functions, B-cells and T-cells (lymphocytes) assume the specific response of the immune system.

3.2. The immune system in action

The immune system acts as a whole in the destruction of microorganisms. Figure 4 illustrates the main steps followed by the immune system during the destruction of a foreign antigen [9]:

1. The subset of cells denominated APC (*Antigen Presenting Cells*), like macrophages, seek and destroy the foreign antigens, fragmenting them in antigenic peptides.
2. Some of the peptides bind to special proteins, called MHC (*Major Histocompatibility Complex*), being presented in the cell surface as a pair "MHC/peptide".
3. The T-cells have receptors in its surface that bind to several of these pairs. This binding process stimulates and activates the T-cells.
4. After the activation, the T-cells release some chemicals (lymphokine) that will activate other components of the immune system, such as B-cells.
5. When activated, B-cells will be divided and differentiated in plasmocytes that will produce a high rate of special molecules, called antibodies. The antibodies produced have the shape of B-cell receptors.
6. The binding between the antibodies produced and antigens will neutralize the intruder.

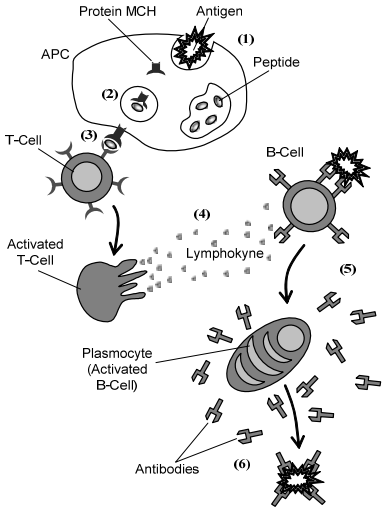

Figure 4 - The immune system activity.

Implicit in the natural immune system functionality are several concepts and positive practical characteristics, such as *adaptability*, *diversity*, *memorization* and *learning*. To achieve this, the immune system possesses several mechanisms necessary for an efficient response, which are described in the following section, where we give further insights into the usefulness and applicability of the AIS approach in our proposal.

3.3. The adaptive immune response applied to computer science

The adaptive immune system has numerous interesting characteristics that caught the interest of computer science researchers, not only in computer security but in other fields, such as robotic, control and pattern recognition [9]. In this section we emphasize the main mechanisms implemented by the immune system and describe the general metaphors involved in a possible adaptation and inclusion into computer security architectures.

First, the meaning of *diversity* is related to the large variety of *B-cells* and *T-cells* receptors that can possibly bind to all types of antigens. The cell activation is only done if the cells receptors have a level of affinity above a defined threshold. In networks, a potential attack should be identified by a specific receptor that should react through adequate security countermeasures. Therefore, the system must have a large diversity of detectors to detect as much attacks as possible. In order to have a low number of false positives, only attacks that are above a defined threshold should be able to activate the detectors.

The immune system can also differentiate between self and non-self cells by the means of a *negative selection process*. This mechanism allows the body to be safe from self cells and prevents it to start being attacked by its own immune system, which would lead to an autoimmune disease. In IDS this mechanism can be used to distinguish normal activity that should be allowed, from abnormal activity that should be considered a possible intrusion. For example, an open TCP port can define a characteristic of the system, accepting connections in that port. Otherwise, all connections to TCP ports not defined in the open state should be considered as potential intrusions.

Considering that a cell can only bind to a specific antigen, it is necessary to have numerous cells to efficiently cope with an infection. The theory of *clonal selection* [10] explains why in each individual, antibodies are only produced for the antigens that he has been previously exposed to. After being stimulated by an antigen, the cell is cloned into a multitude of copies. This process creates a huge group of cells capable of attacking that specific antigen (see Figure 5). In the case of B-cells, the clonal selection process produces two kinds of cells: the plasmocytes and memory B-cells. The former will be responsible for a large scale production of antibodies to fight the antigens. The later are long life cells that can remain in the body for years, whose main function is to react faster to a second attack of a similar specific antigen.

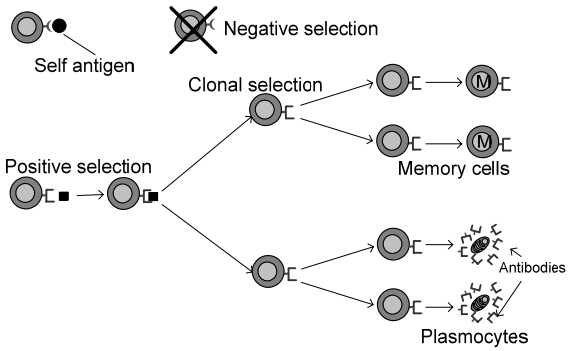

Figure 5 - Clonal and negative selection in B-cells.

In a distributed environment there are several sources of events that should be considered in the evaluation of a potential attack. For example, various IDS sensors (cells) should be installed in the network and the host events should be collected for analysis. The IDS sensors can thus be seen as cells that work together in a distributed way, collecting and analyzing events for further reaction. Finally, the memory cells should be generated by the intrusion detection system to guarantee a future faster reaction to similar attacks.

An AIS is thus an abstract model that captures some aspects of the human immune system and can be used to experiment and predict some of its activities. It is a recent and emergent field of research which tries to define immunological models that can be used in several areas, such as anomaly and intrusion detection [14][18][24].

4. Related work

The use of immune system metaphors has inspired research in numerous fields, computer security being one of the most promising and popular. The idea of bringing the natural biological defence mechanism into the computational world is very appealing and in fact, within the realm of computer security, there are

currently two main areas of considerable interest in the applicability and development of artificial immune systems, namely: the eradication and spread control of computer viruses and worms and the detection of network intrusions. In this section we describe the major developments done so far in immunity-inspired intrusion detection systems for computer networks, stressing some of the main characteristics of the developed prototypes.

4.1. AIS applied to IDS

The seminal work of Forrest *et al.* [36] took advantage of some features of the immune systems that are desirable for the unpredictable, uncontrolled and open environment in which computers currently operate. These characteristics include diversity, adaptability, anomaly detection, multiple layers and identity by behaviour, among others. The authors also argue that these experiments and ideas can open the way for a much wider variety of computer immune systems architectures. This work resulted in LISYS [5][20], an artificial immune system for network intrusion detection, based on biological immune system mechanisms.

In [19] Forrest proposed a first approach to an AIS in the area of network security where the non-self is characterized 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 counterpart of the artificial immune system 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, this time with a much reduced activation threshold.

In [25] and [26] Kim and Bentley identified three fundamental design goals requirements for the derivation of network based intrusion detection systems: *distribution*, *self-organization* and *lightweight* operation. The AIS frameworks for network intrusion detection should also include three distinct and self-organized evolutionary algorithms: negative selection, clonal selection and gene library evolution. In [24] Kim presents an AIS that incorporates the requirements and characteristics listed above, describes the developed architecture and shows results of its application in a real local area network.

Dasgupta [16] proposed an agent-based framework for intrusion/anomaly detection and reaction in networked computers. The mobile agents are able to interact with each other by travelling around the network nodes and monitoring several parameters, such as type of user and its privileges, amount of free memory and type of connection. Other Dasgupta's contributions in computer security subject can be found in [13] and [15].

In [3] Aickelin *et al.* present a review of the intrusion detection systems based on AIS developed thus far, stressing their weaknesses and defending the need to adopt a new paradigm, the Danger Theory, introduced in the next section.

4.2. Danger Theory

The Danger Theory [41] is a new paradigm that is gaining increasing popularity amongst immunologists. This theory is not yet completed and is currently surrounded by a lot of controversy [42].

In classical theoretical immunology, the immune response is triggered when the body encounters something that is non-self or foreign, in a discrimination process known as self-non-self recognition. Unfortunately this process is not yet fully understood. There are some natural phenomenons that cannot be explained by the well known classical immunological theories. Theoretical immunologists believe that the difference between "self" and "non-self" is learnt during the maturation process, through the elimination of the T and the B cells that react to the self, in a process know as self elimination [27].

Matzinger's Danger Theory [30][41] 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 the successful transplants where there are no attacks against foreign (non-self) tissues.

A central idea of the Danger Theory is that the immune system does not react to non-self but to *danger*. The system discriminates "some" self and "some" non-self, which starts 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 the 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 new theory suggests that the immune system reaction to threats is based on the correlation of various signals, providing a method of linking thread directly to the attacker.

Aickelin *et al.* [1] aims to investigate the correlation described above and transpose the danger theory (DT) to the realms of computer security. In his approach the self-non-self discrimination is still used but no longer essential, since the reaction will be based on *danger signals*. He proposed an AIS based on DT ideas that is capable of handling the IDS alert correlation problems described above.

5. Proposed Framework

In this section we present a new framework for intrusion detection systems, which is an extension of the CIDF, detailed in [12] and [23]. This new approach applies immunological concepts to a definition of normal traffic in the network.

The CIDF was the result of an effort to develop tools and application programming interfaces (API) so that intrusion detection research projects could evolve from a common reference and modular architecture. Some of the CIDF presuppositions have been further explored by the IDWG [22] of IETF [21]. The purpose of the IDWG was to define data formats and exchange procedures for sharing information of interest between IDS and management systems. The CIDF models an IDS as an aggregate of four components or boxes with specific requirements and roles [17][23] (see Figure 6): the event collector (E-box), the event analysis (A-box), the event database (D-box) and the response unit (R-box). These components are logical entities that interoperate by processing, storing and signalling events.

Different IDS can be implemented by modifying the way events are detected and classified, through the methods used by the E-box and A-box.

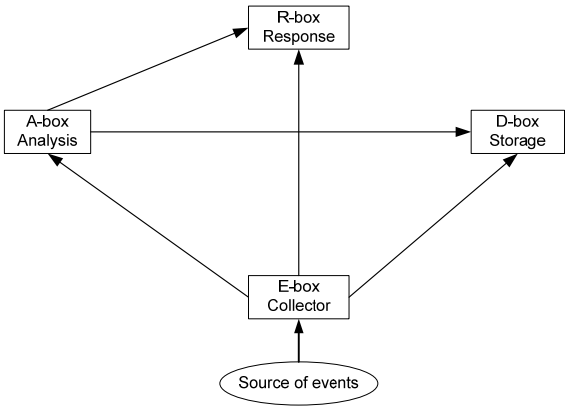

Figure 6 - Main components of CIDF.

Figure 7 shows a block diagram of the I_2NF (***I**mmunity-**I**nspired **N**etwork Intrusion Detection System **F**ramework*) model, which is basically the CIDF model plus an I-box, concerning the new adaptive function and some links to support required extra relations. The boxes inherited from the CIDF model have a similar function, as well as its relations. However, to better understand the proposed model, a brief description of each box is desirable.

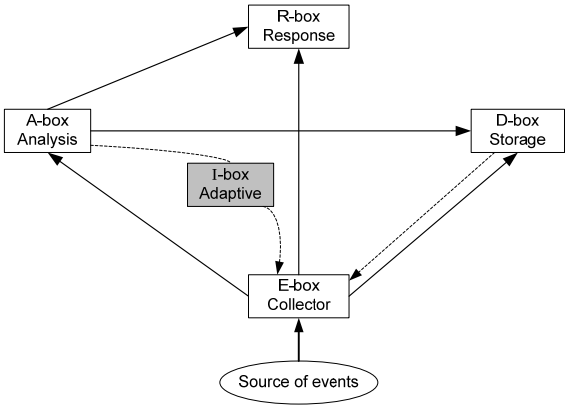

Figure 7 - I_2NF architecture.

This way, the boxes' main functions are:

- *Event collector* (*E-box*). This component is basically a sensor that captures and classify events, based on the audit data produced by the computational environment, such as network packets (datagrams) and/or hosts and applications log files (users by session, number of process, system files changes, etc.). These events are then transmitted to the event analysis component (A-box), to the event database unit

(D-box), mainly for forensic analysis, and/or directly to the response box (R-box). The E-box generates raw or low level events.

- *Event analysis* (*A-box*). The role of this component is to analyze the events received from the E-box. This unit produces high level interpreted events, such as alarms (high level description of the malicious activity) that can be used to react accordingly (R-box) and/or store events (D-box) for later analysis. After a fusion process, the events are correlated and interpreted. The fusion process homogenizes the events, since they arrive from several sources and miscellaneous generators. The event correlation enables the pattern detection from event atomic sets, in order to find out signs of intrusions, attacks, anomalies or policy violations.

- *Adaptive analysis* (*I-box*). The events generated by the A-box can also be used to generate new attack profiles, based on adaptive methodologies. This new component (I-box) should use immune algorithms to generate new event profiles (basically new cells) for the E-box, allowing the system to "learn" and better respond to future malicious attacks. This approach allows the IDS to "grow up" in an adaptive and evolutionary way, being self-adjusted by previously learned attacks.

- *Event database* (*D-box*). All the events collected by the E-box and produced by A-box are stored on the event database (D-box), guaranteeing persistence and allowing a subsequent analysis for later retrieval. This box keeps the memory of past attacks (memory cells), crucial for the production of new event profiles (new cells). These profiles, related to new learned attacks, are dynamically included in the sensors (E-box), allowing a faster response to similar events. This concept of memory and generation of new profiles is similar to the function of memory cells in the immune system, since both intend to react faster to new instances of previously learned attacks.

- *Response unit* (*R-box*). This component, also called countermeasure unit, receives low level events from the E-box and high level events from the A-box, disclosing real or signs of attempted intrusions. In response it applies countermeasures according to the alarms generated. Typical activities may include information actions (printing reports, setting alarms and dispatching emails), defensive actions (killing processes and modifying firewall settings) and survival actions (resetting network connections and activation of alternative systems).

The proposed framework has two main innovations for CIDF. First, the definition and classification of what should be considered "normal" based on the information collected by the sensor. This classification provides a better definition of what should be considered a normal behaviour, in contrast to the signature of "abnormal" (or misuse) implemented by the more conventional IDS.

Secondly, after the classification of normal traffic, there are two different ways, based on immunological concepts, to adapt the sensor with new events profiles: through the I-box and by the learning mechanism of past events that have been stored in D-box.

6. Conclusions and future work

In this document we present a new framework for intrusion detection based on CIDF, which implements some aspects inspired by the biological immune system. We must stress that this deployment has the potential to solve some of the well known problems already identified in current IDS misuse and anomaly detection methods. Furthermore, an immunological approach can heavily contribute to the development of self-adjusted, adaptive and evolutionary intrusion detection systems.

This framework innovates in the definition of normal behaviour in the network and in the use of immune-based concepts and algorithms to guarantee the maintenance of normality. We think that the memory and adaptability concepts, inspired in the immune system, provide the capacity to learn from new attacks and optimize the response time for further occurrences of similar incidents.

Our work will include the development of a prototype to test this framework. We also intend to develop a "plug in" to the Snort IDS [35] implementing these concepts.

7. Acknowledgements

We'd like to thank the contribution and fruitful discussions we had with Professor Henrique Santos and his PhD student Rui Monteiro, both from the

University of Minho, during the writing of this document.

8. References

[1] Aickelin U., Bentley P., Cayzer S., Kim J., and McLeod J., "Danger Theory: The Link between AIS and IDS?", Proceedings of the 2nd International Conference on Artificial Immune Systems (ICARIS 2003) Edinburgh, 2003, pp. 147-155

[2] Aickelin U., Cayzer S., "The Danger Theory and its application to Artificial Immune Systems", In proceedings of the 1st International Conference on Artificial Immune Systems (ICARIS 2002), Canterbury UK, 2002, pp. 141-148

[3] Aickelin U., Greensmith J. and Twycross J.; "Immune System Approaches to Intrusion Detection - A Review"; Proceedings of the 2nd International Conference on Artificial Immune Systems (ICARIS 2003), 2003, pp. 316-329

[4] Bai Y. and Kobayashi H., "Intrusion Detection Systems: Technology and Development", Proceedings of 17th International Conference on Advanced Information Networking and Applications (AINA 2003), China, 2003, pp. 710-715

[5] Balthrop J., Esponda F., Forrest S. and Glickman M., "Coverage and Generalization in an Artificial Immune System". Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2002), 2002, pp. 3-10

[6] Bass T, "Intrusion detection systems and multisensor data fusion", Communications of the ACM, 2000, vol. 43 no.4, pp. 99-105

[7] Bishop M., "What is Computer Security?", IEEE Security & Privacy Magazine, 2003, vol.1 issue:1 pp. 67-69

[8] Burmester G, Pezzuto A. and Wirth J., *Color Atlas of Immunology*, Thieme Publishing Group, March 2003, ISBN 3131267410

[9] Castro L and Timmis J, *Artificial Immune Systems: A New Computational Intelligence Approach*, Springer-Verlag, Heidelberg - Germany, 2002, ISBN: 1852335947

[10] Castro L, Von Zuben F, "The Clonal Selection Algorithm with Engineering Applications", Workshop Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2000), 2000, pp 36-37

[11] Castro L., *Immune Engineering: Development of Computational Tools Inspired by the Artificial Immune Systems*, PhD Thesis - 2001, DCA FEEC/UNICAMP, Campinas/SP, Brazil

[12] Common Intrusion Detection Framework (CIDF) - http://www.isi.edu/gost/cidf/

[13] Dasgupta D. "An Immune Agent Architecture for Intrusion Detection", Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2000), 2000, pp. 42-44

[14] Dasgupta D. and Attoh-Okine N., "Immunity-based systems: A survey"; In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Orlando, Florida, October 1997, vol. 1, pp. 369-374

[15] Dasgupta D. and Gonzalez F., "An Immunogenetic Approach to Intrusion Detection", CS Technical Report (No. CS-01-001), The University of Memphis, May 2001

[16] Dasgupta D. and González F.; "An immunity-based technique to characterize intrusions in computer networks"; IEEE Transactions on evolutionary computation, 2002, vol. 6, No 3, pp. 281-291

[17] Debar H., Dacier M. and Wepi A., "Towards a taxonomy of intrusion-detection systems", Computer Networks, 1999, vol.31 no. 8, pp. 805-822

[18] Forrest S., Hofmeyr S., "Engineering an immune system"; Graft, 2001, vol.4 no.5, pp.5-9

[19] Forrest S., Perelson A., Allen L. and Cherukuri R., "Self-non-self discrimination In a computer", Proceedings of the IEEE symposium on Research in Security and Privacy, May 1994, pp. 202-212

[20] Hofmeyr S. and Forrest S., "Immunity by Design: An Artificial Immune System", Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 1999), 1999, pp. 1289-1296

[21] Internet Engineer Task Force (IETF) - www.ietf.org

[22] Intrusion Detection Working Group (IDWG) - IETF http://www.ietf.org/html.charters/OLD/idwg-charter.html

[23] Kahn C., Porras P., Staniford-Chen S. and Tung B., "A Common Intrusion Detection Framework", submitted to the Journal of Computer Security, 2000

[24] Kim J, *Integrating Artificial Immune Algorithms for Intrusion Detection*, PhD Thesis – July 2002, Department of Computer Science – University College London, UK

[25] Kim J. and Bentley P., "An Artificial Immune Model for Network Intrusion Detection", 7th European Congress on Intelligent Techniques and Soft Computing (EUFIT99), Aachen, Germany, 1999

[26] Kim J. and Bentley P., "The Human Immune System and Network Intrusion Detection", 7th European Congress on Intelligent Techniques and Soft Computing (EUFIT99), Aachen, Germany, 1999

[27] Kuby J., Goldsby R., Kindt T. and Osborne, B., *Immunology 5th Ed*, W. H. Freeman, 2002, ISBN: 0716749475

[28] Landwehr C., “Computer Security”, International Journal of Information Security (IJIS), Springer Berlin Heidelberg, 2001, vol. 1, pp. 3-13

[29] Lee W., Stolfo S. and Mok K., “A data mining framework for building intrusion detection models”, IEEE Symposium on Security and Privacy, 1999, pp. 0120

[30] Matzinger P., “The Danger Model: A Renewed Sense of Self”, 2002, Science 296, pp. 301-305.

[31] McHugh J., “Intrusion and Intrusion Detection”, International Journal of Information Security (IJIS), Springer Berlin Heidelberg, 2001, vol. 1, no.1 , pp. 14-35

[32] Portnoy L., Eskin E. and Stolfo S., “Intrusion detection with unlabeled data using clustering”, Proceedings of ACM CSS Workshop on Data Mining Applied to Security (DMSA-2001), Philadelphia USA, 2001, pp. 76-105

[33] Santos H., “ISO/IEC 27001 – A norma das normas em segurança da informação”, Revista “Qualidade – Associação Portuguesa para a Qualidade”, 2006, vol. XXXV, Nº1 Primavera, pp. 11-19

[34] Sherif S. and Dearmond T., “Intrusion Detection: Systems and Models”, Proceedings of IEEE Enabling Technologies: Infrastructure for Collaborative Enterprises - WET ICE’02, Pittsburgh USA, 2002, pp. 115-133

[35] Snort - the de facto standard for intrusion detection and prevention; www.snort.org

[36] Somayaji A., Hofmeyr S. and Forrest S., “Principles of a Computer Immune System”; Proceedings of New Security Paradigms Workshop, Langdale, Cumbria, ACM, 1998, pp. 75-82

[37] Taylor C. and Alves-Foss J., “An empirical analysis of NATE: Network Analysis of Anomalous Traffic Events”, Proceedings of New Security paradigms Workshop, ACM Press New York, 2002, pp. 18-26

[38] Taylor C. and Alves-Foss J., “NATE: Network Analysis of Anomalous Traffic Events, a low-cost Approach”, Proceedings of New Security paradigms Workshop, ACM Press New York, 2001

[39] The ISO 17799 Directory - http://www.iso-17799.com

[40] The ISO17799 Toolkit: ISO17799 and ISO27001 - http://www.27000-toolkit.com/index.htm

[41] The real function of the immune system by Polly Matzinger - http://cmmg.biosci.wayne.edu/asg/polly.html

[42] Vance R., “Cutting Edge Commentary: A Copernican Revolution? Doubts About the Danger Theory”, The Journal of Immunology, 2000, no. 165, pp. 1725-1728

[43] Venter H. S., Eloff J., “A Taxonomy for Information Security Technologies”, Computers and Security, 2003, vol.22, number 4, pp. 299-307