Intrusion Prevention System Inspired Immune Systems

Yousef Farhaoui
Department of Computer Science, Faculty of sciences and Techniques, Moulay Ismail University
Errachidia, Morocco
e-mail: youseffarhaoui@gmail.com

Abstract
In view of new communication and information technologies that appeared with the emergence of networks and Internet, the computer security became a major challenge, and works in this research axis are increasingly numerous. Various tools and mechanisms are developed in order to guarantee a safety level up to the requirements of modern life. Among them, intrusion detection and prevention systems (IDPS) intended to locate activities or abnormal behaviors suspect to be detrimental to the correct operation of the system. The purpose of this work is the design and the realization of an IDPS inspired from natural immune systems. The study of biological systems to get inspired from them for the resolution of computer science problems is an axis of the artificial intelligence field which gave rise to robust and effective methods by their natural function, the immune systems aroused the interest of researchers in the intrusion detection field, taking into account the similarities of natural immune system (NIS) and IDPS objectives. Within the framework of this work, we conceived an IDPS inspired from natural immune system and implemented by using a directed approach. A platform was developed and tests were carried out in order to assess our system performances.

Keyword: intrusion prevention system, intrusion detection system, artificial immune system, security systems

Copyright © 2016 Institute of Advanced Engineering and Science. All rights reserved.

1. Introduction
Computer attacks have been since their appearance a real threat. With their great diversity and specificity to systems, these can have catastrophic consequences. Various measures to prevent these attacks or reduce their severity exist but there is no complete solution.

The IPS is one of these currently most effective measures. Their role is to recognize intrusions or attempted intrusions by abnormal user behavior or recognition of attack from the network data stream. Different methods and approaches have been adopted for the design of IPS. Among these methods, one is inspired by nature, especially immune systems [1-3], which have properties and great similarity to IPS.

The study of the immune system is promising new area of research (artificial intelligence), namely, artificial immune systems (AIS) [13]. These are actually modeling, implementation and adaptation of concepts and methods of biological immune systems to solve problems.

As part of our work, we focus on the immune systems for detection and intrusion prevention. Our goal is to develop an artificial immune system for our intrusion prevention system, implementing the main immune theories. To evaluate performance, we will conduct a series of tests to analyze the results in order to measure the contribution of immune systems in the intrusion prevention [6, 7].

Intrusion prevention systems and immune systems are characterized by their hierarchical architecture and their distributed operation on a set of subsystems. To better model these notions, we will adopt a method of designing an IPS.

2. Natural Immune Systems (NIS)

2.1. NIS Properties
The NIS is a source of inspiration for new branches of IT. With very important properties, it becomes a valuable reference. Several research works have been developed on the basis of it functioning.
2.1.1. Discrimination Entre Self and Non-Self
The most important property which is the basis of immune reactions is the ability of the NIS to distinguish between self cells and non-self cells and the ability to recognize the exact type of each foreign cell [6], [7].

2.1.2. Learning and Memory
In each contact with a new kind of antigens, the NIS categorizes it and keeps it in mind, thanks to cell division mechanism followed by a selection process to refine and improve the response of NIS in the next contact with the same antigen. This allows the NIS to increase efficiency for the recognition of antigens; this process is called affinity maturation [8].

2.1.3. Communication and Dissemination
The different actors of NIS need to exchange messages under the form of signals. Two types of dialogues exist: one-way signals which transit by immunological components or continuous dialogues by an exchange of molecular signals [9].

2.2. Immune Theories
The behavior and reactions of the NIS are primarily governed by immune theories.

2.2.1. Negative / Positive Selection Theory
This theory manages the process of creating cells. Specifically, this theory manages the creative process at the level of the discrimination between self and non-self. Lymphocytes have receptors on their surfaces lymphocytes from the bone marrow migrate to the thymus, at this stage they are called immature or naïve T cells. Their para-topes undergo a process of pseudo-random genetic rearrangement, after a very important test is introduced [10].

2.2.2. Clonal Selection Theory
The recognition of an antigen by B cells, they produce specific antibodies. The antibody associate with the antigen using receptor then using cells such as T aideuses, B cells of stimulated and a proliferation process allows B cells to reproduce by creating clones themselves [11]. A second process will select among those new cells with high affinity to make memory cells [12].

3. Artificial Immune Systems (AIS)
The AIS is a new branch of artificial intelligence. It is designed to solve various problems, inspired from remarkable properties and concepts of biological immune system [13]. AIS are a mathematical or computer implementation of the operation of natural immune system.

3.1. Modeling AIS
The common model known by the Framework of AIS, defines the rules to be complied by AIS and the process for developing new approaches. The necessary conditions are [14]:

- The representation of system components.
- Adapting procedures to monitor the evolution of the system. The three conditions mentioned above are imperative for the development of a framework to define AIS [8].

Then, the form of an antibody as a set of l parameters. These parameters may be represented by a point in a space of l dimensions. A first notes that in this plan, those antibodies are close to each other. Population or repertoire of N individuals is modeled as a space forms a finite volume V containing N points. An antigen is represented by the point Ag = <Ag1, Ag2, ... , Agl>, an antibody is also represented by a point Ab = <Ab1, Ab2, ..., Abl>. To measure the degree of completeness between the antigen and the antibodies, several techniques can be used. More often the distances are used [15]:

- Euclidean distance
  \[ D = \sqrt{\sum_{i=1}^{l} (Ab_i - Ag_i)^2} \]

- Manhattan distance
  \[ D = \sum_{i=1}^{l} |Ab_i - Ag_i| \]
Hamming distance

\[ D = \sum_{i=1}^{l} \delta_i \]

with

\[ \delta_i = \begin{cases} 1 & \text{if } A_{bi} \neq A_{gi} \\ 0 & \text{if not} \end{cases} \]

So, we notice that the antigen-antibody affinity is relative to the distance in the space between them. Once the antigens and antibodies are represented, the quantitative function of the defined Completeness degree between them, it remains only to implement the immune theories.

### 3.2. Immune Algorithms

#### 3.2.1. Clonal Selection Algorithm

This theory is based on the principle that only the cells having the antigen recognize the antigen proliferate and become memory cells. The clonal selection algorithm is based on the following:

- Holding a set of memory cells.
- Selection and cloning of the most stimulated antibodies.
- Re-selection clones proportionally to the affinity with the antigen.
- Removal of unstimulated antibodies.

The maturation of their affinity [8].

```plaintext
Begin
    P = set of shapes to be recognized
    M = Population random individuals
    while (A minimal form is not recognized)
        for i de 1 à taille(P)
            aff = affinite(P, M)
        end for
        Select n1 elements having the best affinity with the elements of M
        Generate copies of these elements in proportion to their affinity with the antigen
        Mutate all copies proportionately with their affinity with the forms of the assembly P
        Add mutated individuals in the population M
        Choose n2 of these mutated elements (optimized) as memory
    end while
End
```

Figure 1. Clonal selection algorithm

#### 3.2.2. Negative Selection Algorithm

This concept is very interesting, especially for systems monitoring applications and detection and prevention of abnormal or unusual uses [14]. The problem of protection of computer systems is the learning problem of distinguishing between self and non-self. Rather, they compare the loads detection problem within systems to the process of adverse selection takes place in the thymus [16].
III. IMMUNE SYSTEMS INTRUSION DETECTION AND PREVENTION SYSTEMS (IDPS)

3.3. Characteristics of IDPS

It is important to recall the functions or very important fundamental properties that must satisfy an IDPS and should be listed [1, 2]. After that, we will try to see what is offered in parallel artificial immune systems and make the analogy between All IDPS [3], [5], [18]:
- Robust: The IDPS must have different points of detection and prevention, and should be highly resistant to attacks.
- Configurable: The IDPS must be easily configurable based on each machine on which it will be deployed. The degree of dependence on the operating system must be minimized.
- Expandable: Adding new hosts in all machines must be monitored elementary and the dependence on operating systems should not be an obstacle to this extension.
- Upgradable: It is necessary that the IDPS can face an unexpected increase in the flow of data to be monitored due to an extension of all the constituents’ hosts the IDPS.
- Adaptable: The IDPS must dynamically adapt to changes (hardware or software) within the network in question.
- Effective: The IDPS should be simple and easy to be deployed in order to avoid affecting the hosts and network performance monitoring.
- Distributed: Special attacks can be detected and stopped after analysis of different signals and alarms from different hosts [19]. The IDPS should be able to recover various events from different stations on the network, analyze them and send responses to different stations.
In order to develop an effective IDPS we will try to find the properties mentioned above in an artificial immune system.

3.3.2. Properties of AIS for Detection and Intrusion Prevention

The immune system is capable of protecting the human body surface to bacteria, viruses or any kind of antigens. This fundamental role is mainly based on discrimination between self and non-self. This discrimination is the key process forming an immune response. Whether or not known antigens, the natural immune system can be compared to an anomaly detector with a very small number of false positives and false negatives [4]. The three most important properties of an IDPS were found in the immune systems. The immune systems are [4], [20].

This article talks about the negative selection algorithm. The algorithm proceeds in two phases. The first is to generate a set of sensors and the second is to use these detectors to monitor data by making a comparison. The comparison may be a comparison of the number of common bits [16], [21], [24].

3.3.3. Immune Systems and Immune Algorithms

Once we have found the necessary properties for our IDPS and the choice of using immune systems has been done. It is interesting to have a method for creating algorithms composed of AIS. A comparison of the components of the immune systems and their equivalents in immune algorithms, allows us to easily design the algorithms forming our artificial immune system components.

<table>
<thead>
<tr>
<th>Immune Systems</th>
<th>Immune algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antigen</td>
<td>Problem to resolved</td>
</tr>
<tr>
<td>Antibody</td>
<td>Vectorbetter solutions</td>
</tr>
<tr>
<td>Recognition of antigens</td>
<td>Identifying the Problem</td>
</tr>
<tr>
<td>Production of antibodies from memory cells</td>
<td>Loading previously best solutions found</td>
</tr>
<tr>
<td>Removal of T cells</td>
<td>Elimination of surplus solutions potential</td>
</tr>
<tr>
<td>Proliferation of antibodies</td>
<td>Use of a process for creating exact copies of the solution</td>
</tr>
</tbody>
</table>

By following this process we can develop immune algorithm. This comparison applies to the different problems, we will be interested only in the design of an IDPS inspired immune systems. The table shows a very adapted comparison:

<table>
<thead>
<tr>
<th>Immune Systems</th>
<th>IDPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thymus and bone marrow</td>
<td>Primary IDPS (supervisor)</td>
</tr>
<tr>
<td>Lymphnode</td>
<td>Local Host</td>
</tr>
<tr>
<td>Antibody</td>
<td>Detector</td>
</tr>
<tr>
<td>Antigen</td>
<td>Intrusion</td>
</tr>
<tr>
<td>Self</td>
<td>Normal activity</td>
</tr>
<tr>
<td>Noself</td>
<td>Abnormal activity (suspicious)</td>
</tr>
</tbody>
</table>

Based on this comparison, they proposed AIS for detection and intrusion prevention. These AIS consists of a primary IDPS which acts as a supervisor and a plurality of second IDPS will be installed on each host in the network. The functioning of this IDPS model is as follows:

3.4.4. Generating Detections

These two points are crucial in creating a detector. Once the elements constituting the detector were listed with the type of each of them, the last step will be to define the values of each detector element as follows. If the item is continuous type, it will be represented by an interval defined by two terminals. Once the elements and their respective values have been listed, the detector will be represented by a data structure containing these elements [17], [23].
3.4.5. Anomaly Detection and Detection by Scenario

We have seen that for the behavioral detection, it is favorable to use the negative selection theory. However, for the second approach (detection per scenario) and which is based on a set of signatures, we will use the clonal selection theory as follows: In the approach of detection by scenario, we have a database containing the set of known attack signatures. Based on these signatures we will generate detectors allowing, after packet analysis, to detect the presence of certain signatures in order to conclude that an intrusion or intrusion attempt has occurred. The choice of the clonal selection theory for scenario approach has been made because in this process, this theory is used to generate and refine antibody for the detection of known antigens. We could compare the clonal selection theory, antibodies and antigens detectors known to attack signatures. So to conclude this is the most frequent use of immune theories for the design of intrusion detection systems; NIDPS with detection by scenario: Theory of clonal selection HIDPS with behavioral detection: Theory of negative selection.

4. Solution Description and Global Architecture of the IDPS Results

We opted for the design of a hybrid IDPS composed of an NIDPS based on the approach of analysis by scenario, implementing the theory of clonal selection and using a signature database and a HIDPS based on behavioral approach, implementing the theory of negative selection and using a user profile database. Using immune theories, the core of our IDPS generates some varied signatures of attacks and user profiles in a pseudorandom manner. This methodology allows us to develop the analyzer to possibly discover new attacks or variants of attacks.

Our IDPS is composed of:
- NIDPS: generating sensors on the basis of signatures. These detectors will be used to analyze network traffic.
- HIDPS: Based on profiles of normal user behavior in order to generate detectors able to recognize unusual behaviors of users.
- Administration Console: From this console, the administrator can configure the different parameters of the IDPS, see the different alerts, start learning control.

The components of our solution to be deployed in this way: The NIDPS will be installed on the machine that is the network proxy to analyze all network packets. While, HIDPS be deployed on all machines that constitute the LAN.

Here the overall architecture of our solution:

![Figure 4. Global Solution diagram](image-url)
5. Databases Used
A large amount of information is analyzed and generated by the various components of our IDPS; whether user profiles, the alerts by the various detectors or a list of attack signatures. The use of databases is very important in the architecture of our IDPS; we opted for the use of three databases:

5.1. The Database "Profiles"
This database contains all information about user profiles. The data contained in the database are generated by the HIDPS during the learning phase. For security reasons, user profiles must pass through the HIDPS supervisor to ensure compliance and consistency of the data in the profile.

5.2. The Database "Signatures"
This data source is very important; it is the basis of NIDPS. It includes all the known attacks using a certain format. The format of the signature is important insofar as all detectors adopt this format. Unfortunately, there is no standard model for the codification of signatures. The signature must represent a reliable, unambiguous and accurate attributes that can recognize the attack. We must remember that the signatures will be used to analyze network traffic. The attributes used to represent an attack should be based on the information in the packets. We can analyze network traffic to multiple levels of granularity. Indeed, we can consider the traffic of a packet perspective, sessions or connections. It is necessary to define the set of attributes to be used from the set of existing attributes [22]. We propose in this paper a particular model of signatures. Our signature model was designed to meet the requirements by an attack signature. The attack signature must represent unambiguously the attack and should only contain information that allows recognizing the attack. In our case, the signatures are coded so as to be modifiable and can model the new attacks, with new analytical methods ... etc. The analysis and synthesis of various network attacks has allowed to classify these into three classes:
- Attacks 'data': These are all recognized attacks by analyzing the data portion of the packet, such as SQL injection attacks. These will be recognized if the following channels ('- -', or 1 = 1) is found in the packet.
- Attacks 'Headers': These are all recognized attacks by analyzing packet headers, such as DOS attacks with spoofing headers.
- Attacks 'Requests/queries': The requests generally include several packages. Some attacks will be recognized by analyzing the set of packets that make up the request, such as attacks of input validation or buffer overflow attacks, which cannot be recognized, that the length or the number of parameters which constitute the request.

In the modeling of different classes of existing attacks, our Signature contains the following fields:

<table>
<thead>
<tr>
<th>Id</th>
<th>Type</th>
<th>Action</th>
<th>Data</th>
<th>Val</th>
<th>Flag</th>
</tr>
</thead>
</table>

- Id: unique identifier of the signature.
- Type: header, data, queries/Request.
- Action: The Action Analysis (eg find a sub string, count the number of attributes, length of a query requested service ... etc.)
- Data: In the case of attributes kind of strings: the desired string.
- Val: In the case of attributes to numeric values: the value of the attribute.
- Flag: Additional information

The identifier serves as an index in the signatures database while the type allows to find the table that contains the signature. The action defined the processing to be used, this is the most important field for a signature, it contains a keyword that shows which method known for analyzing data. Various actions have been implemented such as:
- Substr: Search for a sub string, this keyword is used much more on the attacks data and queries.
- LenStr: Calculate the length of a string, retrieves attacks such as DOS attacks.
- ValidStr: This shows whether the character strings do not contain invalid characters.

The 'Data', if you look for a string (eg The SUBSTR action) contains the string in which to search. The 'Val', in case the action returns a numeric value contains the numeric value that...
can say that this is an attack. The 'Flag' is an optional field serving parameters for the analytical method. At this signature model, we will assign an interpreter to run the action of analyzing each signature. At any time if we want to increase the number of signatures by adding new, just use the previously defined model, and if you need new analytical functions, we added them to the interpreter.

5.3. The Database "Alerts"
This database will list all alerts generated by the detectors of the two components of IDPS (NIDPS and HIDPS). An alert should inform the administrator about suspicious event, providing enough information: time, date, sensor, signature or abnormal behavior, the attacker, the victim. This database will be accessed by the administrator to identify traces of attacks or anomalous behavior.

6. HIDPS with Behavioral Approach
The first stage of deployment HIDPS, is undoubtedly the learning step, during which it traces back to normal user behavior by creating a profile for each. User profiles are a source of data that can tell us about the behavior of users. We chose to use the following information to model a user profile:
- Name of the user.
- Root directory.
- Average consumption CPU and RAM
- Opening time / closing sessions.
Other information could have been used, such as the average consumption of bandwidth, most visited websites, the response speed to the operating system messages.

6.1. Architecture HIDPS
Our HIDPS will consist of a HIDPS supervisor and a plurality of HIDPS slaves to be deployed throughout the network components machinery. The theory of negative selection is the HIDPS core. This theory runs in two phases: generation of detectors and attack prevention and behavior analysis. The first phase runs on the HIDPS supervisor, who sends alarms generated at HIDPS slaves to execute the second phase of the theory. This consists of analyzing the actual behavior of the user on the basis of sensors.

6.2. HIDPS Supervisor
HIDPS the supervisor's role is to:
- Extract the users of the database profiles.
- Generate detectors and send them to HIDPS slaves by running the first phase of the theory of negative selection those generating sensors that gather all the necessary information for the analysis of user behavior in the future.
- Analyze the HIDPS of reports slaves and list alerts in a database.
- Send commands to start the learning phases, analysis, launch and stopping HIDPS slaves.

6.3. HIDPS Slaves
The main role of HIDPS slaves role is to:
- Generate user profiles during the learning phase.
- Use event sensors to extract the current behavior of the user.
- Run the second phase of the theory of negative selection, which involves using sensors generated by the first phase in order to analyze the behavior of the user.

6.4. Theory of the Negative Selection
Our HIDPS is based on this theory; it can generate alarms from the user profile, and set up at the end to recognize suspicious behavior. As we have previously seen this theory runs in two phases:
Phase I: Generation of detections
This stage runs on the HIDPS supervisor. During this phase, we extract user profiles from the database. Each profile will be considered the self system, and will be used for the
random generation of detectors. Then, a test is set up to purge all alarms generated by keeping only those who do not recognize the self-chain.

Figure 5. Phase I of the negative selection (generation of detections)

Phase II: Analysis

This phase runs on HIDPS slaves. During this phase, we operate the detectors generated by the preceding phase to conduct the analysis of the current behavior of the user. The HIDPS slave must have sensors to inform him about the current behavior of the user. A function will measure the degree of resemblance between that conduct and detectors previously generated and an alert is generated if it reaches a certain percentage....'

Figure 6. Phase II of negative selection (Analysis)

6.5. Operation HIDPS

The HIDPS are deploying and starting in two phases:
- Learning phase: The HIDPS supervisor sends the command from the beginning of the learning phase for different HIDS slaves. During the learning phase, the HIDPS slave periodically retrieves user behavior information. For numeric values, the HIDPS slave calculates the average of different values extracted. The profile generated by each HIDPS slave will then be sent to HIDPS supervisor, who is in charge of the list.
- Monitoring Phase: During this phase, the supervisor HIDPS extract the profiles of each user, applies the first phase of the negative selection theory to generate detectors. Detectors will be sent to each slave HIDPS with the start command of the monitoring phase.
7. NIDPS with Scenario Approach

The second important component is the NIDPS using analysis with scenario approach. This approach requires a database of known attack signatures on the basis of these signatures, the core of NIDPS generates detectors, can recognize the original signature, but also the signatures derived from the latter. The NIDPS core contains mainly the analysis function; it is based on the theory of clonal selection. The function analysis of our NIDPS contains both detectors generating process and their introduction to the packet-flow analysis.

7.1. Architecture NIDPS

a. Manager

This is the manager of the solution. The manager is responsible for:
- Starting the different components.
- Assigning different analysis tasks.
- Extracting attack signatures and generating detectors, performing clonal selection algorithm.
- Receive reports and list alerts.

b. Sensor

The sensor is responsible for capturing network packets. Different 'sensors' can be deployed in our solution to make this lighter task. If one opts for the deployment of several 'sensors', you must define for each the subset of network traffic that will capture (eg TCP, UDP ... etc.).

c. Analyzer

The analyzer is actually comparable to an antibody which is tasked to monitor and recognize certain types of antigens. In our case, the antigen in question is the attack signature to recognize. So the analyzer receives the signatures of the 'Manager' and puts in place to recognize a type of attack. We opted for the joint use of 'Analyzers Sensors'. This use guarantees a lighter and autonomous solution.

7.2. Operation NIDPS

Our analysis uses NIDPS with scenario approach, based on the theory of clonal selection; it uses as a source of data network packets. Here are the steps for its implementation:

- Packet Capture: The first step of the analysis is capturing packets, through the 'sensors' that capture and transmit network packets to 'analyzers' to conduct analysis. At this level, you can also save the captured packets in data structures to analyze them later if the administrator chooses to defer analysis.
- Extraction and formatting attributes: This step allows you to extract a high level of attribute vector from the captured packets to be analyzed later. This step is very important; it helps to prepare the packages for the analysis phase by making some changes on them.
attribute analysis: once the 'manager' has generated a set of detectors by applying the theory of clonal selection, the analysis function performed by the analyzer 'compares to the type of signature, a set detectors with the attributes of packets. Based on this comparison, many reports are generated.

figure 11. mode of operation nids

8. conclusion

the objective of this work was to design and implement an idps inspired for immune systems. the idps is a very important brick in a security system, several research studies using different methods and approaches are devoted to them. among them, artificial immune systems, inspired by the natural immune systems, can be very interesting for the field of intrusion detection, given the similarity of features and objectives of the latter. we focused on the two main theories that are the basis of the immune response, namely the theory of clonal selection and the theory of negative selection. the study of these two theories immune in the case of intrusion detection showed that the theory of clonal selection is more appropriate for the scenario analysis, while the theory of negative selection is more appropriate to behavioral analysis. the choice of implementing an idps is very important, especially if one considers that the idps will be deployed on a network with multiple machines with different hardware and software configurations. the idps is designed hierarchically and is distributed across multiple machines and requires the analysis of data from different sources. so we designed a hybrid idps (nidps + hidps), analyzing the two sources of information and using both immune theories. tests on our solution aimed to define the contribution of immune systems for intrusion detection. the use of clonal theory can generate from an attack signature more detectors can recognize not only the attack in question but also variants of this attack, or further similar attacks. however, the use of the theory of negative selection in the case of analysis with behavioral approach is to detect any abnormal behavior and which is different from the typical behavior of the user. the test analysis allowed inferring that the application of theories is beneficial for immune recognition of new forms of attack. indeed, these methods, since they make use of random processes in the generation phase detector, generate a large number of false positives.

references


