Immunological Approach for Intrusion Detection

Meriem Zekri, Labiba Souici-Meslati

Special issue CARI'12

Immunological Approach for Intrusion Detection

Meriem Zekri – Labiba Souici-Meslati

Laboratoire LISCO

Département d'Informatique
Université Badji Mokhtar – Annaba
BP 12, 23000 Annaba
ALGERIE

zekri_meriem@yahoo.fr, labiba.souici@univ-annaba.org

RÉSUMÉ. L'un des défis centraux en sécurité informatique est de pouvoir déterminer la différence entre un comportement normal et un comportement potentiellement dangereux d'un système. Pendant des décennies, les développeurs ont protégé leurs systèmes en utilisant des méthodes classiques. Cependant, la croissance et la complexité des systèmes informatiques ou de réseaux à protéger nécessitent le développement d'outils de défense automatisés et adaptatifs. Des solutions prometteuses voient le jour avec l'informatique inspirée de la biologie, et, en particulier, l'approche immunologique. Dans cet article, nous proposons deux systèmes immunitaires artificiels pour la détection d'intrusion en utilisant la base de données *KDD Cup'99*. Le premier est basé sur la théorie du danger en utilisant l'algorithme des cellules dendritiques et le second est basé sur la sélection négative. Les résultats obtenus sont prometteurs.

ABSTRACT. One of the central challenges with computer security is determining the difference between normal and potentially harmful behavior. For decades, developers have protected their systems using classical methods. However, the growth and complexity of computer systems or networks to protect require the development of automated and adaptive defensive tools. Promising solutions are emerging with biological inspired computing, and in particular, the immunological approach. In this paper, we propose two artificial immune systems for intrusion detection using the *KDD Cup'99* database. The first one is based on the danger theory using the dendritic cells algorithm and the second is based on negative selection. The obtained results are promising.

MOTS-CLÉS : Systèmes immunitaires artificiels, Détection d'intrusion, Détection d'anomalies, Théorie du danger, Algorithme des cellules dendritiques, Algorithme de la sélection négative.

KEYWORDS: Artificial immune system, Intrusion detection, Anomaly detection, Danger theory, Dentritic cell algorithm, Negative selection algorithm.

1. Introduction

The network security of computer systems and networks is very important and motivates many researches to find solutions. Intrusion detection is one of those solutions that detect intrusion of unwanted users. The challenge is to create a system able to differentiate between normal, non-offensive to the system, and harmful use. However, this challenge is not easy to overcome; computer systems and network to protect are becoming more complex and data to deal with is increasing. In addition, the common types of intrusions lead us to develop automatic and adaptive defense tools. Intrusion detection tools use several techniques to help them determine what qualifies an intrusion versus normal traffic.

An intrusion detection system uses anomaly detection or misuse detection, our study focuses on the anomaly detection which involves discrimination between normal and abnormal data, based on normal data knowledge. Compared to a more traditional approach, anomaly detection has the clear advantage of detecting new intrusions. Several systems have been designed to solve the problem of intrusion detection, but many of them may be subject to the generation of false alarms. In recent years, a recent bio-inspired paradigm started to prove its ability in many areas, such as pattern recognition and data mining. This paradigm corresponds to artificial immune systems (AIS) inspired by the natural immune systems [5]. Its effectiveness has encouraged researchers to study and learn from the immune mechanisms for the implementation of artificial systems that can effectively detect intrusions [16].There are several models based on theoretical models of the immune system. We are particularly interested by the danger theory (DT), which had a tumultuous beginning caused by several doubts about many of its concept. However, a few years ago, a group of British researchers has extensively studied the DT, they even called their project "The Danger project" [3]. The danger theory [1] involves two basic algorithms that are the dendritic cell algorithm (DCA) and Tolk-like Receptor (TLR). The DCA algorithm was developed to detect anomalies; therefore, it seems most appropriate for our work, besides the fact that it is an algorithm of the danger theory which greatly interested us since the beginning of our work on artificial immune systems because this theory corresponds to a relatively new concept in natural immunology. Indeed, while most models are based on immunological discrimination self / non-self where all foreign bodies are detected and removed, the danger theory, in turn, is based on the detection of danger and not the detection of strangeness. Recent research on the DCA algorithm [11, 12, 13] show that it has not only promising performance of the detection rate, but it can also help in reducing the number of false alarms, compared to similar systems.

The aim of our work is to design two artificial immune systems for intrusion classification: the first is based on the dentritic cell algorithm (DCA) while the second is

based on the negative selection algorithm (NSA) which is one of the first immune models proposed for intrusion detection. We compare the performance of these two immune approaches to determine which is most appropriate for the given problem, using the well-known KDD cup'99 dataset.

Our paper is organized as follows: In the second section, we introduce the artificial immune systems, followed by the intrusion detection systems in the third section. In the fourth section, we discuss the application of artificial immune systems in the intrusion detection field. In the fifth section, we summarize major works using the immunological approach for intrusion detection. A presentation of the chosen artificial immune algorithms is followed by a description of the dataset, experiments and results in sections six to nine. At the end of this article, we give our conclusions and prospects for future extensions.

2. Artificial Immune Systems (AIS)

Artificial immune systems represent a class of algorithms inspired by the principles and functioning of the innate immune system. These algorithms typically exploit the characteristics of the biological immune systems in terms of learning and memory as means of solving complex problems [25].

Immunological Concepts and entities	**Immunity based models**	**Computer problems**
Self/no Self: T cells recognition.	Negative selection algorithm	Errors, anomaly detection and change.
Idiotypics networks, immunological memory, B cell.	Immune networks theory	Supervised and unsupervised learning
clonal Expansion, maturation, B cell	Clonal selection. algorithm	Search and optimization
Innate Immunity	Danger theory	defense strategy

Table1. Computational immunity based models and specific immunological concepts [5].

The field of artificial immunology has evolved gradually since 1985, with growing interest towards the development of computational models inspired by several immunological principles. Some models mimic the abstract mechanisms of biological immune system to better understand its natural processes and simulate its dynamic behavior in the presence of antigens or pathogens while others focus on the design of

algorithms, using simplification techniques (sometimes outdated) of various immunological processes [5]. The central principle of immunology is that the immune system responds to the presence of foreign entities (called non-Self) and does not respond to the host (called the Self). Table 1 summarizes the main immunity-based models and their corresponding concepts, entities and applications.

The study of the danger theory considers two aspects of the hazard model. The immunologists examine potential danger signals and how to be affected cells of the immune system. In collaboration with immunologists, computer scientists have sought ways to model the formation of the danger that could be used in the improvement of AIS. This is done to improve the anomalies detection systems for computers on networks. There are two developed algorithms inspired by the danger theory, the Tolk-like Receptor algorithm (Twycross, 2007) and the dendritic cells algorithm (Greensmith, 2006) [11].

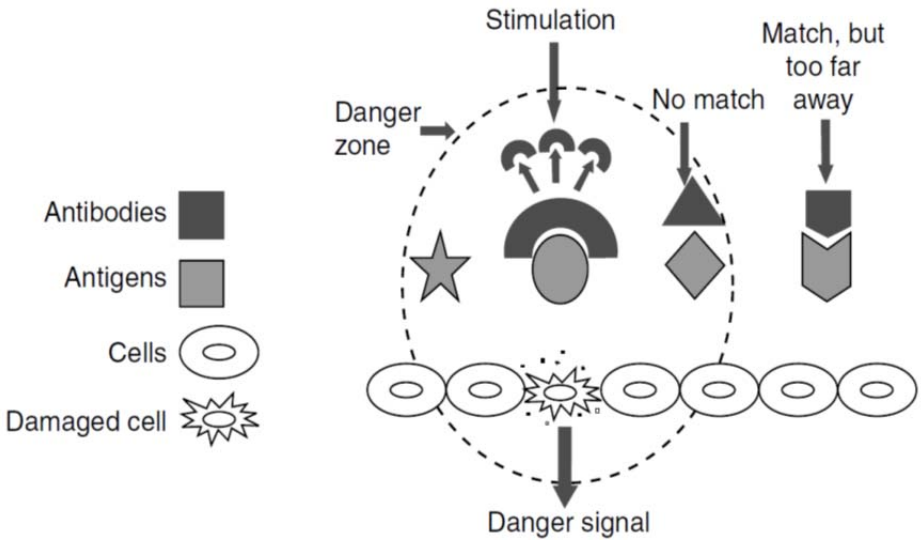

Figure1. Illustration of the danger theory [5]

The Tolk-like Receptor algorithm (TLR) [2] is based on two populations of interacting cells, dendritic cells (DCs) and T-cells. The DCs implemented in TLR collect antigen from an antigen store, and process signals. Unlike the DCA, different

categories of input signals are not used, with focusing on the nature of the interactions between DCs and T-cells. In TLR, DCs are created as immature detectors which sample signals and antigens for a finite specified period of time. If the DC receives a signal during antigen collection, it is termed mature, and conversely, DCs which did not detect the presence of a signal are termed semi-mature.

The dendritic cells algorithm (DCA) is based on an abstract model of the behavior of dendritic cells (DC). In nature, the DC perform the antigen presentation function, where the debris found in the tissues is collected by DC, then processed to form the antigen and present it to the adaptive immune system in combination with context information. This information is obtained by processing the signals of the different developing DCs found in the tissue at the time of antigen collection. As a technique for calculating, DCA performs the correlation of the context, derived by processing a set of input signals, with the antigen and the correlated data [12].

3. Intrusion Detection

In computer security, intrusion detection is the act of detecting actions that attempt to compromise the confidentiality, integrity or availability of a resource. Intrusion detection can be performed manually or automatically. In the process of manual intrusion detection, a human analyst is reviewing the log files to look for suspicious signs that might indicate an intrusion [6]. A system that performs automated intrusion detection system is known as intrusion detection system (IDS). An IDS is characterized by the detection method, the behavior of the detection, location of the source audit and frequency of use.

There are two main approaches to detect intrusions. The first consists in searching for known signatures of attacks, while the second aims to define normal behavior of the system and research what does not fit into this behavior. An intrusion detection system by searching signatures knows what is wrong while an intrusion detection system by analyzing behavior knows what is good. We are therefore talking about malware detection and anomaly detection.

Malware detection operates essentially by looking for abusive activities, compared with abstract descriptions of what is considered malicious. This approach attempts to get in shape the rules that describe the unintended uses, based on past intrusions or known theoretical weaknesses. The rules can be made to recognize a single event that is itself dangerous to the system or a sequence of events representing an intrusion scenario. The effectiveness of this detection is based on acuity and coverage of all possible abuses by the rules [21].

Anomaly detection technique behavior is quite old, it is also used to detect suspicious behavior in telephony. The main idea is to model, during a learning period,

the “normal” behavior of a system/program/user, defining a line of conduct (profile), and then considering (detection phase) as suspect any unusual behavior (significant deviations compared to the model of "normal" behavior) [20].

The behavior after the detection describes the response of the IDS to attacks. When responding to attacks by taking either proactive or corrective actions, the IDS is called active. If the IDS simply generate alarms, it is called passive. The location of the source audit distinguishes between the IDS based on the type of input information they analyze [21]. This input information can be the audit paths, system logs or network packets. The frequency of use is an orthogonal concept, some IDS capabilities have continuous monitoring in real time, while others must be performed periodically. The first three characteristics are grouped into the functional category, because they concern the internal operation of the intrusion detection engine, namely its input information, the reasoning mechanism and its interaction with the information system. The fourth characteristic distinguishes the RTID (Real-Time Intrusion Detection) from scanners used for security evaluation.

4. Immunological approaches for Intrusion Detection

The use of AIS for intrusion detection is an interesting concept for two reasons: the immune system provides protection in a distributed manner against intruders as well as being adaptive and classical techniques used in computer security are unable to cope with the dynamic and increasingly complex computer systems and security.

To provide viable IDS, AIS must construct a set of sensors that accurately measure the corresponding antigens. In the current immunological approaches dedicated to IDS, network connections and sensors are modeled as strings. Detectors are randomly created and then undergo a maturation phase. If the sensors do not match any of this, they are eliminated; otherwise they become mature [16]. These mature detectors start to check new connections during their lifetime.

If these mature detectors do not match anything else, exceeding a certain threshold, they are activated. This is then reported to a human operator who decides if there is a real anomaly. Such an approach is known as negative selection, since only the detectors (antibodies) do not survive. However, this approach shows attractive problems of scaling when applied to real network traffic. Hence, other immunological approaches have been applied to the IDS, such as clonal selection, immune networks and the theory of risk. The danger theory has provided the most promising results, particularly the dendritic cell algorithm (DCA, Dendritic Cell Algorithm) whose main ability is to manage large data.

5. Summary of Some Major Works on AIS for IDS

Since the early nineties until today, several artificial immune systems were proposed to detect different types of intrusions on a computer network. This research increased with the emergence of Internet and the explosion of attempted theft of confidential or non-confidential data, fraud, and exploitation of resources of others...

Similar works to ours have been performed in recent years, particularly those of Julie Greensmith [11, 12, 13] on the DCA algorithm and its application to anomaly detection, there were also comparisons of the DCA algorithm with the TLR algorithm [2] and Kohonen self-organizing maps (SOM) [14], as well as the adequacy study of NSA for the problems of intrusion detection [26]. The results of these works are encouraging and put the DCA algorithm as the most appropriate approach for the intrusion detection problem.

A very recent work [17] exploiting the KDD'cup 99 database to experiment a modified version of the DCA algorithm to compare it with the classic version of the algorithm, has provided very good results proving that the DCA algorithm still remains one of more suitable for solving the intrusion detection problem.

Year	Ref.	AIS Models	ID Approach	Dataset
2005	[10]	Danger theory Dendritic cell algorithm	Anomaly Detection	Breast cancer standard dataset.
	[4]	Negative Selection	Network based intrusion detection	KDD'cup 99 dataset
2006	[26]	Negative Selection vs. positive Selection	Anomaly Detection and network intrusion detection.	Biomedical dataset and Iris-Fisher.
	[23]	Artificial immune system: generation of detectors	Anomaly Detection	KDD'cup 99 dataset
2007	[2]	Innate immunity, theory danger: TLR and Dendritic cell algorithm	Anomaly Detection	Libtissue API[1]
	[18]	Cooperative AIS	Malware detection and network based detection	KDD'cup 99 dataset

[1] A prototype software system for the construction of the second generation of AIS and apply them to real world problems.

2008	[14]	Comparison between Dendritic cell algorithm and the Self Organizing Map	Malware detection	Data: a block of 254 IP addresses for each of the 70 guests.
	[19]	The innate immune system	Malicious Code Detection	
2009	[30]	Principle of antigen-antibody reactions	Intrusion detection	DARPA intrusion detection dataset
	[8]	Negative selection	Intrusion detection	Sendmail, Lpr and STIDE Process data sets
2010	[9]	Neural network combined with artificial immune system	Malware detection and Network-based IDS: detection of malware code	Network worms, Trojans, classic viruses. (according to the Kaspersky classification)
	[24]	Clonal selection	Detection of malware processes	API call sequences
2011	[29]	Immune Intrusion Detection Algorithm	Intrusion detection	KDD'cup 99 dataset
	[27]	Artificial Immune	Harmful Information Filtering	600 pornographic and 600 nonpornographic Web pages
2012	[20]	Negative Selection	Intrusion detection	arbitrary set of data
	[7]	Dendritic cell algorithm	Malware detection	
2013	[22]	Dendritic cell algorithm	Intrusion Detection for Wireless Sensor Networks	Input signals : emitting one application message per second
	[17]	Negative Selection and Dendritic cell algorithm	Intrusion detection	KDD'cup 99 dataset

Table 2. Summary of some major works on AIS for IDS

6. The Dendritic Cell Algorithm (DCA)

The dendritic cell algorithm (DCA) is a correlation algorithm that can perform anomaly detection on classified data sets. The merger process of the signal is inspired by the interaction between dendritic cells (DCs) and their environment. The DCA has the ability to combine multiple signals to assess the current context of the environment. The correlation between the context and the antigen is used as the basis of anomaly detection in this algorithm [5].

The main components of the algorithm DC are:

1) Individual dendritic cells with the ability to perform a multi-signal processing.
2) The collection and presentation of antigens.
3) Sampling behavior and state changes.
4) The population of DCs and their interactions with signals and antigens.
5) Signals and antigens incoming with pre-classified signals.
6) Presentation of multiple antigens and analysis using the type of antigen.
7) Generation abnormality coefficient for different types of antigens.

The dendritic cell is a signal processing unit, which takes a binary decision (yes / no) whether the antigen it has collected during its life, was collected under normal conditions or not.

The antigens are required; they represent the data to be classified with the basis of the classification that does not follow from the structure of these antigens, but the relative proportions of the three categories of input signals which are: “PAMP”, “danger” and “safe” [14] (see Table 3):

– PAMP: indicates the presence of definite anomaly.

– Danger Signal (DS): may or may not indicate the presence of anomaly, but the probability of being anomalous is increasing as the value increases.

– Safe Signal (SS): indicates the presence of absolute normal.

Signal	*Biological property*	*Computational example*
PAMP	Indicator of the microbes presence	Error message per second
DS	Indicator of tissue damage	Network packet per second
SS	Indicator of healthy tissue	Size of network packets

Table 3. Signal functions in DCA.

The output signals of the DCA process associated with predefined weights to produce three output signals. The three output signals are the co-stimulatory signal (CSM), the semi-mature signal (Semi) and mature signal (Mat). Predefined weights

used are presented in Table 4 and the equation for calculating output signals is as follows:

$$O_j = \sum_{i=0}^{2} (W_{ij} \times S_i) \quad \forall j \qquad (1)$$

	PAMP S_0	**Danger signal** S_1	**Safe signal** S_2
Csm O_0	2	1	2
Semi O_1	0	0	2
Mat O_2	2	1	-2

Table 4. Suggested weights for Equation (1)

Where O_j are the output signals, S_i are the input signals and W_{ij} is the transforming weight from S_i to O_j .

Algorithm 1. Pseudo code of DCA.

Inputs: S= input signals pre-categorized + antigens. / **Outputs:** E=antigens + MCAV (Mature Context Antigen Value).

- Create an initial population of dendritic cells (DCs), D
- Randomly select 10 DCs from DC population;

For *each selected DC* **Do**
- Get the antigen;
- Store the antigen;
- Get the signals;
- Calculate interim output signals;
- Update the cumulative output signals;

If *cumulative Csm > migration threshold* **Then**
- Remove the DC population;
- Assign the cell-context to DC;

If *cumulative Semi<=cumulative Mat* **Then**
Cell context=1;
Else
Cell context=0;
End
- All DCs which collected the antigen and have a cell-context out for analysis;
- Terminate this DC and add a naive DC to the population;

Else
- DC back to population;

End
For each incoming data **Do**
- Calculate the number of mature DC and semi-mature DC;

If *nb semi-mature DC > nb* mature *DC* **Then**
Antigen = normal;
MCAV = 0
Else
Antigen = abnormal;
MCAV = 1;
End

End

The DCA introduces the migration thresholds assigned individually to determine the life of a DC. This can make the algorithm sufficiently robust and flexible to detect the antigens found during certain periods. The individual sums the DC output signals resulting in cumulative Csm, cumulative Semi and cumulative Mat. This process continues until the cell reaches the end of its useful life, that is, the cumulative Csm exceeds the migration threshold; the DC stops sampling signals and antigens. At this point, the other two cumulative signals are assessed. If the cumulative Semi is greater than the cumulative Mat value, the cell differentiates towards semi-mature state and is assigned a 'context value' of 0, and vice versa, greater cumulative Mat results in the differentiation towards mature state and a 'context value' of 1 [14]. To assess the

potentially anomalous nature of an antigen, a coefficient is derived from the total values of the population, called MCAV (Mature Context Antigen Value) of this antigen.

This is the proportion of mature presentation of context (context value of 1) of this particular antigen, compared to the total amount of antigens presented. This result in a value between 0 and 1 for which a threshold anomaly, called "Threshold MCAV" can be applied. The reported value for this threshold reflects normal or abnormal items presented in the initial set of data. Once this value has been applied, the antigens with MCAV that exceeds this threshold are classified as abnormal and vice versa.

7. The Negative Selection Algorithm (NSA)

The negative selection algorithm is the first artificial immune algorithm that has been proposed for intrusion detection. The intrusion detection process of NSA consists of three main phases; (1) the definition of self, (2) generation of detectors and (3) monitoring of occurrence of anomalies (see Figure 2). There are two ways to implement the negative selection algorithm: with V-detectors (variable number of detectors) and with C-detectors (constant number of detectors) [26], which has been chosen in our work

Algorithm 2. Pseudo code of NSA

Input: $S \subseteq U \equiv$ labeled data "normal"**,** l, r where l: string length and r matching threshold ;

Output: detectors set $D \subseteq U$;

Begin

- Generate a set (D) of detectors (such that each fails to match any element in S);
- Monitor new sample $\delta \in U$ (by continually checking the detectors in D against δ ;

If any detector matches **Then**

-Classify as normal;

Else

-Classify as abnormal;

End

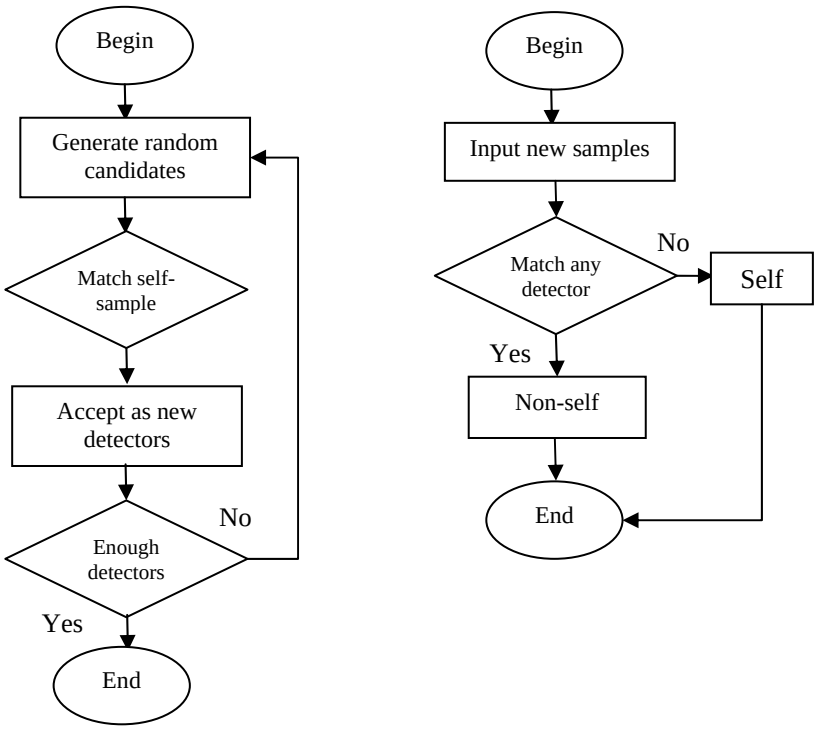

Figure 2. Detectors generation process (left) and Monitoring process (right)

8. The dataset and Preprocessing

The KDD cup'99 dataset is derived from the DARPA 98, the Lincoln Laboratory data set for the application of data mining techniques in the field of intrusion detection. *KDD cup'99* summarizes the two sources of data connections (data instances), each connection has 41 attributes. *KDD Cup'99* is one of the few available labeled data sets in the field of intrusion detection. Instances of data connections are labeled as normal or attacks types. As intrusion detection systems by artificial immune assumes the existence of two classes, the labels of each instance of data in the original data set are replaced by either "normal" for normal connections or "abnormal" for attacks. Because of the abundance of attributes, it is necessary to reduce the size of the data set by removing the irrelevant attributes. For this, the information gain is calculated for each attribute and attributes with lowest information gain are removed from the data set [28].

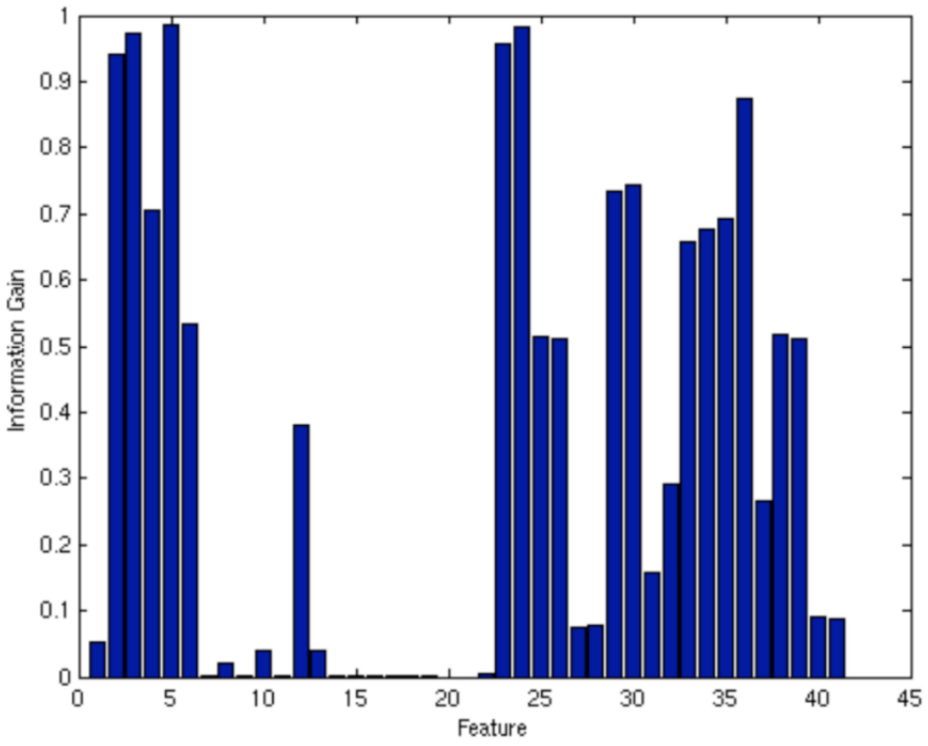

Figure 3. Information gain of each attribute [15]

The information gain of an attribute indicates the statistical significance of this attribute relative to the classification [15]. The information gain, called *gain (S, A)* of an attribute *A* with respect to a collection of examples *S*, is defined in the following equation:

$$Gain\ (S, A) \equiv Entropy\ (s) - \sum_{v \in Values(A)} (\frac{|S_v|}{|S|} Entropy(S_v)) \qquad (2)$$

Where Values (A) corresponds to all possible values of the attribute A, and is the subset of S for which attribute A to a value v, the equation of entropy is (the proportion of S belonging to class *i*):

$$Entropy\ (S) = \sum_{i=1}^{2} -p_i\ \log_2 p_i \qquad (3)$$

After calculating the information gain of 10% of the database KDD cup'99, we get the histogram shown in Figure 3.

9. Experiments and Results

Our experiments consist in the implementation of two artificial immune algorithms, which are the dendritic cell algorithm (DCA) and the negative selection algorithm (NSA) with C-detectors. Both algorithms were implemented in Java in NetBeans IDE. ROC analysis (receiver operating characteristic) is performed to evaluate the performance of the classification of the DCA and the NSA. If the example is non-self and is classified as non-self, it is counted as a true positive but if it is classified as self, it is counted as a false negative. If the example is self and it is classified as self, it is counted as a true negative but if it is classified as non-self, it is counted as a false positive (see Figure 4).

The rate of true positives (TP), false negative (FN), false positives (FP) and true negatives (TN) of each experiment are calculated in addition to the detection rate (DR) and the false alarms rate (FAR).

		Actual	
		P	**N**
Classifier	**P**	TP	FP
Predict	**N**	FN	TN

Table 5. ROC analysis confusion matrix [26]

$$Detection\ rate\ (DT) = \frac{TP}{TP + FN} \qquad (4)$$

$$False\ alarm\ rate\ (FAR) = \frac{FP}{TN + FP} \quad (5)$$

For all experiments concerning the DCA algorithm, the size of the dendritic cells population was set at 100 and remains constant in all iterations of the system. The threshold for dendritic cell migration for each individual is chosen randomly between 100 and 300 to ensure the survival of the cell after several iterations of the system.

We applied some variations in the implementation of two algorithms, they are described as follows:

- **Experiment 1:** DCA with a continuous data loading.
- **Experiment 2:** DCA with a random data loading.
- **Experiment 3:** NSA with a random loading of 1000 detectors with different values of *r* (2, 3, 4, 5, 6).
- **Experiment 4:** NSA with a random loading of a single detector.

We have run each program several times, 10 iterations for DCA and 100 iterations for the NSA (only 10 iterations for the DCA since its execution time is relatively larger than the NSA, from 15 to 20 min for iteration and needs a lot of memory, against just 30 sec for the NSA, which requires very little storage space). We used ROC analysis [26] to measure the actual performance of our classifiers. We wanted to test if the order of the data could affect the proper working of the DCA. The results of the first two experiments indicate a slight decrease in detection rate when the data are randomly selected. We tried to change the weight for calculating output signals, the result of this change was a disaster, and no data were correctly classified.

DCA provided good performance in terms of false alarm rate, which is 0; this means that one of the objectives of the anomaly detection has been achieved because it is important that there are the least possible false alarms. We also noted that, when the data set is small (1000 records for example), the classification of the DCA is excellent and the true positive rate is relatively high (0.99 or 1.00).

For the NSA algorithm, we also made a random loading of 1000 detectors and single detector, with which the correspondence took place with all of our examples. The use of more than a randomly selected detector provides better results than the use of one. Another variant of the NSA algorithm is the change in the value of *r* (*r* contiguous bits matching rule), which has greatly affected the classification.

Category		TP	TN	FP	FN	DR	FAR
Experiment 1		0.7154	1.00	0.00	0.2846	0.7154	0.00
Experiment 2		0.6521	1.00	0.00	0.3179	0.6821	0.00
Experiment 3	*r = 2*	0.9211	0.4294	0.3705	0.0799	0.9211	0.4631
	r = 3	0.7548	0.5183	0.2361	0.2452	0.7548	0.3129
	r = 4	0.3455	0.6324	0.2005	0.6545	0.3455	0.2407
	r = 5	0.2845	0.7128	0.0085	0.7155	0.2845	0.0102
	r = 6	0.0814	0.1985	0.0007	0.9186	0.0814	0.0035
Experiment 4		0.7121	0.4987	0.2147	0.2879	0.1210	0.3009

Table 6. The ROC results of our experiments

We obtained very variable results, when r = 6, there is a hardly correct classification, and the results improve gradually as r decreases. Therefore, the value of r seems very important, smaller is r, better is the classification performance. This seems obvious, as it does the matching between two attributes, which is insufficient to judge the proper functioning of the system. Note also that greater is r, worse is the classification performance.

We have encountered another problem with the NSA algorithm, especially when the choice of a single detector is made: the results are truly random, during the run of 10 consecutive iterations, we can get a real positive rate ranging from 0.08 to 0.25 and 0.50 in some cases, making the NSA algorithm unstable and we cannot rely on its results.

Unlike the DCA algorithm, which has raised no false alarms, the NSA issued a large number, making it unreliable and inadequate for the anomaly detection. Thus, compared to the NSA, DCA correctly handles large data sets and gives satisfactory and promising results.

10. Conclusion and future work

We used two algorithms in the field of the immunological detection of anomalies with the *KDD cup'99* dataset. The results for the dendritic cell algorithm (DCA) are quite encouraging and show that we can further improve the implementation of this algorithm to obtain better results. In contrast, the negative selection algorithm (NSA), did not provide conclusive results, it emits a large number of false alarms in contrast to the DCA algorithm whose false alarm rate is around zero. We also note that NSA has difficulty in managing a large data set, which is a serious drawback, given the current size of database computer systems.

Artificial immune systems (AIS) are promising solutions in the field of intrusion detection. Research around these systems is still the focus of several researchers, in order to exploit all the concepts and mechanisms for identification and detection used by the innate immune system.

Future researches that can be applied to the DCA algorithm are to find a way to make it more adaptive and flexible. We can also try to test with different data sets and also to make rigorous comparisons with other immunological approaches to see where it stands in relation to the performance of other methods

More generally, it would be interesting to conduct further comparisons between immunological classifiers and other ones, which can be bio-inspired or not, considering interesting applications such as intrusion detection. These comparative studies will certainly lead researchers to very interesting conclusions in the attractive field of bio-inspired computing...

11. Bibliography

[1] Aickelin U., Bentley P., Cayzer S, Kim J., McLeod J. Danger Theory: The Link between AIS and IDS?, *2nd International Conference on Artificial Immune Systems*, Edinburgh, U.K. September, 2003

[2] Aickelin U., Greensmith J., Sensing Danger: Innate Immunology for Intrusion Detection, *Elsevier Information Security Technical Reports*, Vol. 12, No. 4, pp. 218-227, 2007.

[3] http://ima.ac.uk/danger "The Danger Project"

[4] Dasgupta D., Gonzalez F. Artificial Immune Systems in Intrusion Detection, chapter 7 of "Enhancing computer security with smart technology" V. Rao Vemuri 2005

[5] Dasgupta D., Nino L. F., *Immunological Computation, theory and application*, Auerbach, 2009

[6] Debar H. An Introduction to Intrusion-Detection Systems, Proceedings of Connect, 2000

[7] Fu J., Yang H. Introducing Adjuvants to Dendritic Cell Algorithm for Stealthy Malware Detection, Fifth International Symposium on Computational Intelligence and Design, Vol. 2, pp. 18 – 22, 2012

[8] Geng L., Jia H. Smart Intrusion Detection Method using Negative Selection Algorithm based on Maximum Entropy Model, In proceeding of International Conference on Artificial Intelligence and Computational Intelligence, Vol. 1, pp. 339– 344, 2009.

[9] Golovko V., Bezobrazov, S., Kachurka, P., Vaitsekhovich, L. Neural Network and Artificial Immune Systems for Malware and Network Intrusion Detection, Advances in Machine Learning II, SCI 263, pp. 485–513, 2010.

[10] Greensmith J., Aickelin, U., Cayzer S. "Introducing Dendritic Cells as a Novel Immune-Inspired Algorithm for Anomaly Detection" *In Proceedings of the 4th international conference on Artificial Immune Systems*, pp. 153-167, 2005

[11] Greensmith J., Twycross J., Aickelin U. Dendritic Cells for Anomaly Detection, Evolutionary Computation, pp. 664–671, 2006

[12] Greensmith J. The Dendritic Cell Algorithm, PhD Thesis, University of Nottingham, 2007

[13] Greensmith J., Aickelin U., DCA for SYN Scan Detection, In proceeding of *Genetic and Evolutionary Computation Conference (GECCO)*, pp. 49–56, 2007

[14] Greensmith J., Feyereisl J., Aickelin U. *The DCA:* SOMe Comparison A comparative study between two biologically-inspired algorithms, *Evolutionary Intelligence*, pp. 85-112, 2008

[15] Kayacik, H. G., Zincir-Heywood, A. N., & Heywood, M. I. Selecting features for intrusion detection: A feature relevance analysis on kdd 99 intrusion detection datasets. In Proceedings of the Third Annual Conference on Privacy, Security and Trust, 2005

[16] Kim J., Bentley P.J., Aickelin U., Grennsmith J., Tedesco G., Twycross J. Immune system approaches to intrusion detection – a review. International journal on Natural Computing Vol. 6, Iss.4, pp. 413– 466, 2007

[17] Kumari K., Jain A., Jain A. An Efficient Approach to Categorize Data Using Improved Dendritic Cell Algorithm with Dempster Belief Theory, In Kumar V., Bhatele M. (eds.), Proceedings of All India Seminar on Biomedical Engineering, Lecture Notes in Bioengineering, pp. 165-172, 2013

[18] Luther K., Bye R., Alpcan T., Muller A., Albayrak S. A Cooperative AIS Framework for Intrusion Detection. In IEEE International Conference on Communications, pp. 1409– 1416, 2007.

[19] Marhusin M. F., Cornforth D., Larkin H. Malicious Code Detection Architecture Inspired by Human Immune System, In Ninth International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, pp. 312 - 317 2008.

[20] Mostardinha P., Faria B.F., Zúquete A., Vistulo de Abreu F. A Negative Selection Approach to Intrusion Detection. In Coello C.A.C., Greensmith J., Krasnogor N., Liò P., Nicosia G., Pavone M. (Eds.) Artificial Immune Systems, Lecture Notes in Computer Science 7597, pp. 178-190, 2012

[21] Pieprzyk J., Hardjono T., Seberry J. Intrusion Detection, In Fundamentals of Computer Security, Book Chapter, 2003

[22] Salmon H.M., de Farias C.M., Loureiro P., Pirmez L., Rossetto S., de A. Rodrigues P.H., Pirmez R., Delicato F.C., da Costa Carmo L.F. R. Intrusion Detection System for Wireless Sensor Networks Using Danger Theory Immune-Inspired Techniques, International Journal of Wireless Information Networks, Vol.20, Iss.1, pp. 39-66, 2013

[23] Seredynski F., Bouvry P. Anomaly detection in TCP/IP networks using immune systems paradigm. *Computer Communications*, vol. 30, pp. 740–749, 2006.

[24] Sheshtawi, K.A., Abdul-Kader H.M., Ismail N.A. Artificial Immune Clonal Selection Classification Algorithms for Classifying Malware and Benign Processes Using API Call Sequences. *International Journal of Computer Science and Network Security*, Vol.10 no.4, 2010.

[25] Simon M. Garrett, How do we evaluate artificial immune systems, *Evolutionary Computation*, Vol. 13, No. 2, pp. 145-178, 2005.

[26] Stibor T. On the Appropriateness of Negative Selection for Anomaly Detection and Network Intrusion Detection, PhD Thesis, Germany, 2006

[27] Sun Y., Zhou X. Artificial Immune for Harmful Information Filtering, In M. Ma (Ed.): Communication Systems and Information Technology, Lecture Notes in Electrical Engineering 100, pp. 125–131, 2011.

[28] Tavallaee M., Bagheri E., Lu W., Ghorbani Ali A., A Detailed Analysis of the KDD CUP 99 Data Set, Proceedings of the Second IEEE international conference on Computational intelligence for security and defense applications, pp. 53-58, 2009.

[29] Wang X. Research of Immune Intrusion Detection Algorithm Based on Semi-supervised Clustering, In Deng H. Miao D., Lei J. (Eds.) Artificial Intelligence and Computational Intelligence Lecture Notes in Computer Science 7003, pp. 69–74, 2011.

[30] Zeng J., Li T., Li G., Li H. A New Intrusion Detection Method Based on Antibody Concentration, Emerging Intelligent Computing Technology and Applications. With Aspects of Artificial Intelligence, Lecture Notes in Computer Science 5755, pp. 500-509, 2009