The Artificial Immune System for Network Intrusion Detection: An Investigation of Clonal Selection with a Negative Selection Operator

Jungwon Kim and Peter J. Bentley
Department of Computer Science,
University College London,
Gower Street, London, WC1E 6BT, U.K.
E-mail:{J.Kim, P.Bentley}@cs.ucl.ac.uk

Abstract- This paper explores the use of an artificial immune system (AIS) for network intrusion detection. As one significant component for a complete AIS, *static clonal selection* with a negative selection operator is developed and the system is described in detail. Two important factors, the detector sample size and the antigen sample size, are investigated in order to generate an appropriate mixture of general and specific detectors for learning non-self antigen patterns. By investigating the results of series of experiments, this paper suggests how to choose appropriate detector and antigen sample sizes. These ideal sizes allow the AIS to achieve a good non-self antigen detection rate with a very low rate of self antigen detection. Furthermore, this paper concludes that the embedded negative selection operator plays an important role in the AIS by helping it to maintain a low false positive detection rate.

1 Introduction

The biological immune system is successful at protecting the human body against a vast variety of foreign pathogens (Tizard, 1995). A growing number of computer scientists have carefully studied the success of this competent natural mechanism and proposed computer immune models for solving various problems including fault diagnosis, virus detection, and mortgage fraud detection (Dasgupta, 1998).

Among these various areas, intrusion detection is a vigorous research area where the employment of an artificial immune system (AIS) has been examined (Dasgupta, 1998; Somayaji, et al, 1997). The main goal of intrusion detection is to detect unauthorised use, misuse and abuse of computer systems by both system insiders and external intruders. Currently many network-based intrusion detection systems (IDS's) have been developed using diverse approaches (Mykerjee et al, 1994). Nevertheless, there still remain unresolved problems to build an effective network-based IDS (Kim and Bentley, 1999a). As one approach of providing the solutions of these problems, previous work (Kim and Bentley, 1999a) identified a set of general requirements for a successful network-based IDS and three design goals to satisfy these requirements: being distributed, self-organising and lightweight. In addition, Kim and Bentley (1999a) introduced a number of remarkable features of human immune systems that satisfy these three design goals. It is anticipated that the adoption of these features should help the construction of an effective network-based IDS.

An overall artificial immune model for network intrusion detection presented in (Kim and Bentley, 1999b) consists of three different evolutionary stages: negative selection, clonal selection, and gene library evolution. This model can be differentiated from the previous work performed by Hofmeyr and Forrest (2000), which also developed the AIS for network intrusion detection. While their AIS mainly relies on negative selection to generate immature detectors, Kim and Bentley's model emphasises the integration of three significant components[1]. The previous work (Kim and Bentley, 2000) showed severe scaling problems to cope with a vast amount of network traffic data when only negative selection is applied to a network intrusion detection problem. Although Hofmeyr obtained promising results from the adoption of negative selection for network intrusion detection, Kim and Bentley (2000) argued that his promising results were gained only when the negative selection was employed to a small subset of network traffic data. The random search feature of negative selection led it to fail in the detection of various network intrusions which require the scrutinisation of immense amounts of network traffic data. Thus this approach is only able to detect a limited number of network intrusions.

This paper investigates the use of the niching strategy

[1] Hofmeyr and Forrest's final system employs some other extensions to support the operation of AIS under a real network environment. Even though it may conform to human immune systems more closely, this approach requires excessive computation time to generate the immature detector set, with no guarantee that the initial detectors are useful when they are distributed to other hosts.

provided by a clonal selection algorithm within an AIS. In order to solve the scaling problem of an independent negative selection algorithm, the artificial immune system described in this paper adopts a clonal selection algorithm which embeds a negative selection operator within it. The paper is organised as follows: section 2 briefly describes the AIS for network intrusion detection proposed by Kim and Bentley (1999b) and outlines anomaly detectors and misuse detector which are two important components of IDS's. Section 3 introduces a clonal selection algorithm with a negative selection operator and shows how this is employed for network intrusion detection. Then, in section 4, detailed implementation points including genotypes, phenotypes, genetic operators and fitness functions are provided. Section 5 describes two series of experiments performed for this work and an analysis of the results. Finally, this paper draws conclusions from this work.

2 Artificial Immune Systems for Network Intrusion Detection

While various artificial immune models have been suggested for diverse purposes (Dasgupta, 1998), previous work (Kim and Bentley, 1999a) introduced the salient functions of the human immune system with respect to network intrusion detection. In this work, we view the normal activities of monitored networks as self and their abnormal activities as non-self and design an AIS for distinguishing normal network activities from abnormal network activities.

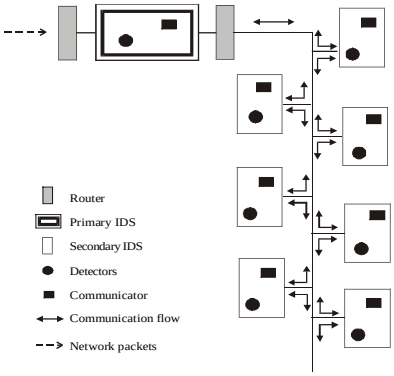

Figure 1 Architecture of the AIS for network intrusion detection.

Based on this view, we proposed a novel AIS for network intrusion detection (Kim and Bentley, 1999b), see figure 1. The AIS for network intrusion detection consists of a primary IDS and secondary IDS's. For the AIS, the primary IDS, which we view as being equivalent to the bone marrow and thymus within the human body, generates numerous detector sets. Each individual detector set describes abnormal patterns of network traffic packets and common patterns of network traffic packets when network intrusion occurs. This unique detector set is transferred to a monitored single local host. We view local hosts as secondary lymph nodes, detectors as antibodies and network intrusions as antigens. At the local hosts (secondary IDS's), detectors are background processes which monitor whether non-self network traffic patterns are observed from network traffic patterns profiled at the monitored local host. The primary IDS and each secondary IDS have communicators to allow the transfer of information between each other, see figure 1.

For the proposed AIS, several sophisticated mechanisms of the human immune system are embedded in three evolutionary stages: gene library evolution, negative selection and clonal selection. These processes allow the AIS to satisfy the identified the main goals for designing effective network-based IDS's (Kim and Bentley, 1999a). This paper focuses on two of these stages: clonal selection and negative selection.

2.1 Anomaly Detection VS Misuse Detection

An IDS is usually comprised of two main components: an anomaly detector and a misuse detector (Mykerjee et al, 1994). The anomaly detector establishes the profiles of normal activities of users, systems, system resources, network traffic and/or services and detects intrusions by identifying significant deviations from the normal behaviour patterns observed from profiles. The misuse detector defines suspicious misuse signatures based on known system vulnerabilities and a security policy. This component probes whether these misuse signatures are present or not in the auditing trails.

One difficulty in developing an effective misuse detector is the creation and update of intrusion signature rules. The work performed in this paper therefore investigates the use of a clonal selection algorithm to provide a more efficient way to build a misuse detector. Clonal selection allows the antibodies of human immune systems to evolve toward existing antigens. This feature is suitable for creating and updating the intrusion signature rules of a misuse detector in an easier way.

3 Related Work

There are many AIS's that have been applied to various fields. Among them, the clonal selection algorithm with negative selection developed for this work is especially motivated by the work performed by Forrest et al (1993) and Smith et al (1993).

Forrest et al (1994; 1997) proposed and used a negative selection algorithm for various anomaly detection problems. This algorithm defines 'self' by building normal behaviour patterns of a monitored system. It generates a number of random patterns that are compared to each self pattern defined. If any randomly generated pattern matches a self pattern, this pattern fails to become a detector and thus it is removed. Otherwise, it becomes a

'detector' pattern and monitors subsequent profiled patterns of the monitored system. During the monitoring stage, if a 'detector' pattern matches any newly profiled pattern, it is then considered that new anomaly must have occurred in the monitored system.

In contrast, Forrest et al (1993) presented the niching strategy of their AIS which follows the analogy of the clonal selection of human immune systems. They explored whether it is able to i) detect common patterns of randomly presented antigens and ii) to discern and maintain the diverse antigen population. In this model, they created one population of antibodies and one population of antigens randomly. They used the GA to evolve the antibody population under a constant antigen population. Conforming to the niching strategy of the human immune system, for each generation, their modified GA selects a random sample of arbitrary size from the antibody population and a single random antigen from the antigen population. After each antibody in the sample is matched against a selected antigen, the fitness score of only one antibody showing the highest match score is increased while the fitness scores of the others remain the same.

Using this algorithm, Forrest et al (1993) showed antibodies evolved to be generalists that match most antigens to some extent. Their analysis of this result showed that antibodies evolved towards finding common schemata that are shared among many antigens. Through various experiments, they observed that this algorithm could sustain multiple different antibody patterns, which appear as multiple peaks in a search space, and the similarity among antigens does not affect this capability. Moreover, they compared this niching strategy of the artificial immune system with the fitness sharing algorithm (Smith et al, 1993). From this comparison, they reported that as the result of the antibody sampling mechanism, the niching strategy of the AIS controls its generality via the antibody sample size. To be more precise, when the sample size decreases, the selective pressures are moved towards generating a population of more general antibodies. Recent work used this algorithm successfully for scheduling (Hart, 1999).

4 A Clonal Selection Algorithm with a Negative Selection Operator

As described in the previous section, this work aims to provide an automated way of building a misuse detector. When network traffic data is gathered under two cases where intrusions are simulated and not simulated, the AIS should generate detectors containing non-self patterns without overlapping self patterns in the data. This is achieved by the clonal selection algorithm, which lets detectors evolve towards the non-self patterns hidden in the collected non-self data.

4.1 Algorithm Description

The AIS for network intrusion detection introduced in this paper adopts the niching strategy of Smith et al's (1993) AIS. Their algorithm used a genetic algorithm to construct the AIS. This work modifies this algorithm to be more appropriate for the network intrusion detection problem. Three major modifications were made to the AIS developed in this work. The first modification is the use of different detector genotype and phenotype representations. Secondly, the fitness and matching functions are altered as the result of detector representation change. Finally, the negative selection stage is embedded in the clonal selection algorithm as an operator. The details of these modifications will be described in the following sections. Figure 2 provides an overview of the system developed during this research.

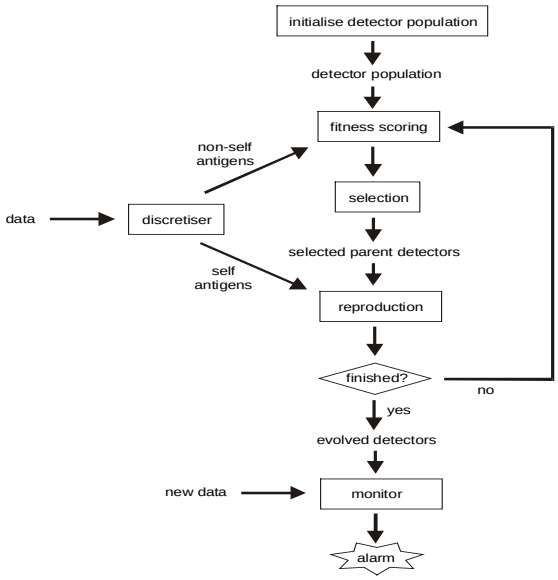

Figure 2 Overview of AIS.

4.2 Providing Self and Non-Self Antigens

As shown in figure 2**,** when the AIS starts, data is fed into the system. In human immune systems, antigens can be divided into two groups: self antigens (our own cells) and non-self antigens (invading pathogens). The clonal selection performed by human immune systems lets antibodies evolve to detect the existing non-self antigens without the detection of any self antigen. The data given to the AIS in this work needs to be divided into a self and a non-self set. Since the clonal selection algorithm employed in this work is used for generating an initial detector set, we assume that the self or non-self class label is already assigned to each antigen data item. In the case when the data has more than two classes, a single class is

predefined as the self and the other classes are regarded as non-self. The self and non-self antigens are then processed by a discretiser before they are passed to the clonal selection module of the AIS.

4.3 Discretiser

The antigen data used in this work consists of a number of attributes. These attributes have continuous and discrete values. Specifically, the continuous attribute values often show a wide range of values. Since the detectors generated in the AIS employs the binary genotypes, a discretisation algorithm is needed. The details of detector genotypes will be discussed in the next section.

There are many discretisation algorithms available and each algorithm has different features (Dougherty et al, 1995). The AIS uses the recursive minimal entropy discretisation algorithm developed by Fayyad and Irani (1993). This algorithm uses the minimal description length theory to minimise the entropy between recursively generated intervals. It improved the classification accuracy of c4.5 and Naive-Bayes algorithms on various data sets and it has been known as one of the best general techniques for a supervised discretisation (Witten and Frank, 2000).

Therefore, the continuous value of an attribute for any antigen data will have been clustered into a number of intervals after the discretiser is applied. The range of each interval and the total number of generated intervals are controlled by the discretisation algorithm.

4.4 Genotypes and Phenotypes

The clonal selection algorithm evolves detectors and these detectors exist as a form of classification rules, which classify non-self from self. A natural expression of classification rules is as a set of disjunctive normal form (DNF) rules. The *if-part* of each rule is a conjunction of one or more conditions to be tested and the *then* side of the rule describes the class label assigned to the rule. In the context of this research, the single detector generated will have a conjunctive rule as its phenotype (Fig 3). Therefore, the universal set of non-self patterns that are detected by the detectors is a disjunction of these conjunctive rules.

The AIS uses simple binary genotypes in order to encode the conjunctive rule detectors. The AIS initialises a detector population by seeding with random genotypes. The detector genotypes consist of a number of genes where each gene represents an attribute of the detector phenotype. The total number of attributes of the given antigen data determines the total number of corresponding genes in the detectors. Each gene is comprised of *nucleotides* and the existing attribute values determine the number of nucleotides. For instance in figure 3, in the case of Attribute 1, its valid values are tcp, udp and imcp. Each nucleotide is a binary bit whose value of one represents the inclusion of the corresponding attribute value in the condition part of a classification rule and whose value of zero indicates the omission of the value (see, figure 3). When all bits are zero, the gene is mapped to a value of NULL.[2] This kind of genotype representation allows a single attribute of each detector rule to have more than one value, which are combined by an "OR" operator. In addition, the existing genes of a detector rule are combined by an "AND" operator.[3]

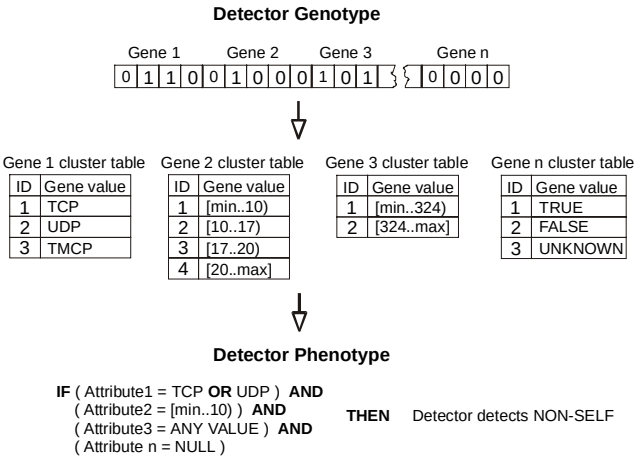

Figure 3 Detector Genotype and Phenotype.

4.5 The Matching Function

Phenotypes mapped from evolved genotypes are represented in the form of detector patterns. As shown in figure 3, an attribute of a detector phenotype is represented by an interval having a lower bound and a higher bound while an attribute of an antigen phenotype is described by one specific value.

Hence, the first step of checking whether a given antigen and a detector match is the comparison of their corresponding attributes. When an antigen attribute value is not within any of the corresponding intervals of a detector phenotype, these two attributes are not matched. For an attribute of nominal type, two genes match when an antigen attribute value is identical to one of the detector phenotype values of its corresponding gene. In order for a given antigen and a detector to match, all the existing genes of the antigen and the detector should match.

---

[2] The first bit of each gene has a special meaning: when it has a value of one, the genotype to phenotype mapping treats the genotype gene as if it is all ones. If it is zero, the remaining bits are used as described. Note that this aspect of the representation was only partially active during tests for vote data, described later, possibly resulting in a slightly degraded TP rate and FP rate. The overall trends were unaffected.

[3] This kind of genotype representation was proposed by De Jong et al (1993) to use the GA for concept learning.

4.6 Fitness Scoring

While the generation of detectors and application of genetic operators are performed at the genotype level, the evaluation of evolved detectors operates at the phenotype level. This is another difference between most work using a negative selection algorithm and clonal selection algorithm (Forrest, et al, 1997; Dasgupta, 1998). Such work usually performed this evaluation procedure on a genotype level using a simple r-contiguous bit matching rule. In contrast, here phenotypes mapped from evolved genotypes are represented in a form of detector rules. These detector phenotypes are evaluated by the following fitness scoring procedure. For a non-self antigen set and its corresponding detector set:

1. *D* detector rules have their fitness values initialised with zeroes.
2. A sample of *D* detector rules is randomly selected from the generated initial *P* detector rules.
3. A sample of *A* non-self antigens are randomly selected from the non-self antigen set.
4. Each detector in the sample is mapped to its phenotype.
5. Each detector phenotype is compared to the selected non-self antigens and the number of matching non-self antigens is counted. This number is defined as a match count for each selected detector.
6. The fitness value of the single detector from the sample that shows the largest match count is increased by the value of the match count. The fitness values of other detectors remain the same. If more than one detector has the largest match count, the fitness value is divided by the number of these tied detectors and their fitness values are increased by the divided fitness value.
7. The processes 2-5 are repeated (for typically three times the number of detectors (Smith et al, 1993)).

As seen in section 3.4, this fitness scoring procedure provides the niching strategy for the AIS. It controls the generality of each detector according to a detector sample size.

4.7 Reproduction and a Negative Selection Operator

After the evaluation of detectors in the detector population, the AIS selects parent detectors for the reproduction of detector offspring. The AIS uses population overlapping where the worst *W*% detectors are replaced by the best *B*% detectors from the newly generated offspring. In addition, a negative selection operator is applied to assure the validity of offspring. This whole reproduction process is described in figure 4.

As shown in figure 4, the offspring detectors are generated by applying crossover and mutations to two parents randomly selected from the fittest *B*% detector rules. The generated offspring are compared to given self antigens. When the offspring matches any self antigen, this offspring is discarded. This kind of invalid offspring can be created because either the parent detectors originally contain some invalid genes or the mutations distort the valid genes of parent detectors. It is not ideal for the AIS to ignore the important and valid genetic information of parents unless it is certain that this kind of bad effect originates from the poor genes of parents. Therefore, when an invalid offspring is produced, the AIS attempts to generate a new offspring by applying the genetic operators to the same pair of parents until the number of failures to generate valid offspring is less than a predefined negative selection threshold, *Nt*. When the number of failures to generate valid offspring is more than *Nt*, the AIS selects a new pair of detector parents and produces new offspring. Offspring generation with negative selection continues until it fills up the empty space of the detector population after the worst *W*% detectors are deleted.

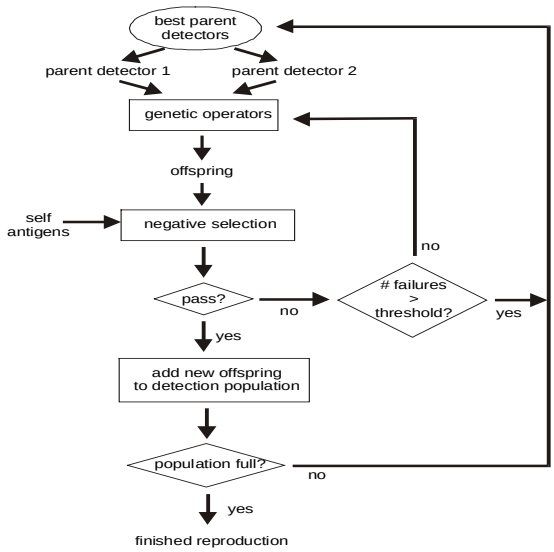

Figure 4 Reproduction and Negative Selection

4.8 Genetic Operators

The clonal selection algorithm presented in this work applies two genetic operators: crossover and mutation. Since a fixed number of nuclotides represents a genotype, a simple one-point crossover is applied by selecting a random crossover point between genes or nucleotides. Furthermore, the following five different types of mutations are introduced:

- Classic mutation: this mutation is a conventional gene flip mutation.
- Generalisation mutation: designed to increase the generality of detectors, it increases the detector generality by causing a new disjunct to be added next to an existing one in the detector phenotype.

- Specialisation mutation: this mutation specialises detectors. This is achieved by causing a random disjunct to be dropped from detector phenotypes.
- Shift Mutation: this shifts all the bits of all the genes to the left or the right direction. The direction to shift is randomly determined.
- Delete Mutation: this mutation flips the first bit of the attribute, changing its corresponding attribute value to 'ANY VALUE' when '1', and back to normal when '0'.

These new mutations are mainly introduced to generalise and specialise detectors. This is because the degree of pattern detection of DNF rules is mainly controlled by doing so.[4]

5 Experiments

This section describes a series of experiments performed to investigate the effects of different detector and antigen sample sizes on the detection rates of the AIS.

5.1 Objective

As introduced in section 3, the detector sample size controls the generality of detectors generated by the clonal selection algorithm. The appropriate mixture of general detectors and specific detectors is critical in order to develop a competent network-based IDS. Detectors should have the maximum level of generality, detecting as many non-self antigen patterns as possible without detecting any self antigen patterns. Furthermore, an ideal detector set should contain detectors showing high specificity that will detect specific antigen patterns found only in a small number of antigens. For these reasons, an ideal detector set should have an appropriate mixture of general detectors and specific detectors. It has been known that the generality of generated detectors is controlled by the detector sample size and the antigen sample size (Forrest et al, 1993; Smith et al, 1993). With these features of AIS in mind, our experiments were performed to understand how best to choose good detector and antigen sample sizes.

5.2 Data and Parameter Setting

This work aims to understand the nature of clonal selection with a negative selection operator. The experiments performed in this paper did not use real network traffic data sets because such sets are typically vast and are not practically suitable for this type of benchmarking work. Instead three different data sets from the UCI repository for machine learning algorithm benchmark work were used (ftp://ftp.ics.uci.edu/pub/machine-learning-databases).

The first data set was Wisconsin breast cancer data. It consists of 699 examples with two classes: 'Malignant' and 'Benign'. 241 examples belong to 'Malignant' and the rest 458 examples belong to 'Benign'. We defined 'Benign' as a self class and 'Malignant' as a non-self class. The detectors generated by the AIS detected 'Malignant' and any data which was not detected by the detectors was regarded as 'Benign'. This set had ten continuous attributes and total 16 missing values. The missing values were filled with random values.

The second data set was the 'vote' data set. This data set is a collection of voting records and each voting record is classified by one of two parties: 'Republican' and 'Democrat'. It consists of 267 democrat and 168 republican examples. Each vote record has 16 voting issues as its attributes and each voting issue has one of three values: yes, no, abstain.

The iris data was used as the final set. It is the most popular data set used in the literature as a pattern recognition test set. It has total of 150 examples with three classes: 'setosa', 'virginia' and 'versicolour'. Each class has 50 examples and every example has four continuous attributes. We prepared three different data sets from this original data set by taking one set as a self set and the rest as a non-self set.

A tenfold cross-validation method was employed to prepare a training set for the AIS to evolve and a test set to detect previously unseen non-self patterns. The tenfold cross-validation method is known as the most robust method from *n*fold cross-validations (Witten and Frank, 2000). A detector population size of 300 was used and best *B*% detector offspring were selected to replace the worst *W*% detectors from parent detectors. 80 was used for both values of *B* and *W*. All mutations occurred with a probability of 0.001 per gene. Each experiment was run for a maximum 50 generations unless it satisfied a termination condition. The termination condition was set as the non-self pattern detection rate for 100% and the self pattern detection for 0%. The threshold of the negative selection operator, *Nt*, was set as 5.

5.3 Experimental Results

Two series of experiments were performed by varying the number of detector sample sizes and the number of antigen sample sizes. Other literature suggests that the generality of detectors is controlled by these two factors (Hart, 1999). The experiments investigated whether the conclusions of the previous work followed our problem: non-self antigen pattern learning from a collected data set.

5.3.1 Varying Detector Sample Size

Table 1 and Table 2 present the results of the first series of experiments, where the number of antigen samples was

[4] These mutations are similar to De Jong's adding and dropping mutations (De Jong et al, 1993).

	Cancer Data			Vote Data		
D	**TP**	**FP**	**TP-FP**	**TP**	**FP**	**TP-FP**
1	93.48 (0.17)	5 (0.26)	88.48 (0.20)	79.43 (0.74)	2.35 (0.09)	77.67 (0.50)
5	94.57 (0.16)	5.83 (0.28)	88.73 (0.36)	88.03 (0.42)	5.29 (0.27)	82.74 (0.84)
10	95.65 (0.12)	5.41(0.58)	90.23 (0.66)	92.49 (0.40)	3.57 (0.25)	88.93 (0.39)
20	95.43 (0.15)	8.33 (0.73)	87.10 (0.52)	94.02 (0.31)	5.29 (0.27)	88.72 (0.47)
30	95.65 (0.13)	6.25 (0.20)	89.40 (0.27)	93.26 (0.33)	5.92 (0.23)	87.34 (0.62)
60	95.87 (0.13)	9.17 (0.53)	86.70 (0.55)	94.40 (0.28)	5.96 (0.15)	88.45 (0.39)
90				95.16 (0.22)	6.65 (0.26)	88.61 (0.57)
240	96.52 (0.097)	10 (0.548)	86.52 (0.7)	95.55 (0.3)	7.13 (0.3)	88.41 (1.07)

Table 1 The mean and variance of true positive rates (TP), false positive rates (FP), and TP-FP rates when an antigen sample size = 1 for various detector sample sizes (*D*). The mean values are followed by the variances in parentheses.

	IRIS Setosa			IRIS Versicolor			IRIS Virginia		
D	**TP**	**FP**	**TP-FP**	**TP**	**FP**	**TP-FP**	**TP**	**FP**	**TP-FP**
1	100 (0)	0.6 (0.036)	99.4 (0.036)	95 (0.011)	4 (8.889E-03)	91 (0.0289)	95 (0.011)	1 (0.0111)	94 (0.044)
5	100 (0)	0.6 (0.036)	99.4 (0.036)	95 (0.011)	4.8 (0.0196)	90.2 (0.0573)	95.8 (0.0036)	0.012 (1.44E-04)	94.8 (0.019)
10	99.8 (4E-03)	1.2 (0.064)	98.6 (0.063)	95 (0.011)	5 (0.0111)	90 (0.0444)	95.6 (7.11E-03)	1 (0.0111)	94.6 (0.0271)
20	100 (0)	0.6 (0.036)	99.4 (0.036)	95 (0.011)	5 (0.0111)	90 (0.044)	95.6 (7.11E-03)	1 (0.0111)	94.6 (0.027)
30	100 (0)	0 (0)	100 (0)	95 (0.011)	5 (0.0111)	90 (0.044)	95.6 (7.11E-03)	1 (0.0111)	94.6 (0.027)
60	100 (0)	0 (0)	100 (0)	95 (0.011)	5 (0.0111)	90 (0.044)	95.8 (4E-03)	1 (0.0111)	94.8 (0.0196)
240	100 (0)	0.6 (0.036)	99.4 (0.036)	95 (0.011)	4.6 (0.0271)	90.4 (0.0693)	95.4 (9.33E-03)	1 (0.0111)	94.4 (0.0338)

Table 2 The mean and variance of TP, FP, TP-FP rates when an antigen sample size = 1 for various detector sample sizes (*D*). The mean values are followed by the variances in parentheses. IRIS class label in each column indicates the assigned self class.

	Cancer Data			Vote Data		
A	**TP**	**FP**	**TP-FP**	**TP**	**FP**	**TP-FP**
1	95.65 (0.12)	5.42 (0.58)	90.23 (0.66)	92.49 (0.40)	3.57 (0.25)	88.93 (0.39)
5	94.35 (0.18)	3.75 (0.40)	90.6 (0.31)	92.14 (0.44)	3.54 (0.07)	88.59 (0.42)
10	95 (0.16)	5.42 (0.39)	89.58 (0.35)	89.56 (0.39)	2.94 (0.17)	86.62 (0.75)
MAX	93.91 (0.24)	5.42 (0.31)	88.5 (0.24)	85.47 (1.63)	3.57 (0.17)	81.90 (2.17)

Table 1 The mean and variance of TP, FP, TP-FP rates when a detector sample size = 10 for various antigen sample sizes (*A*). The mean values are followed by the variances in parentheses.

fixed and the number of detector samples was varied. The detection rate of the system was described by a True Positive (TP) rate and a False Positive (FP) rate. TP was "non-self" detection rate and FP was the rate at which “self” was mistakenly detected by a generated detector set. The desired system should have a high TP and a low FP. The tables show the means and variances of 10 experiments.

For three data sets, the average TP rates generally showed a good level of accuracy, i.e. more than 93%. For the iris data set the best TP rate reached 100%. There were only a couple of cases showing less than a 90% TP rate. The average FP rate was consistently lower than 10% for all cases, but this figure decreased to around 5-6% when *D* was less than 60 for both the cancer and the vote data sets. For the iris data set, the worst FP average rate was only 1%.

As table 1 explains, the TP rate increased as the detector sample size *D* increased. From three data sets, the results of the vote data set showed this tendency most clearly. In order to confirm this result, paired sample t-tests were performed on the vote data results. To find the point at which the difference between TP rates becomes statistically significant, t-tests were performed on the pairs of results and each pair was made by taking two adjacent detector sample sizes. The t-test showed that the difference between the TP rates of $D = 1$ and $D = 5$ was statistically significant with 95% confidence. A two-sided t-test of means produced a p-value of 4.3216%. The t-tests of the rest of pairs produced much larger p-values ranging from 14.7285% to 75.385%. In addition, these p-values became larger as the pair was made from larger sample sizes. These results of the t-tests imply that the difference between the average TP rates with varying detector sample sizes converged as the detector sample size increased. Even though the difference of the TP-rate for different sample sizes was very small for the cancer data, the same kind of tendency was observed. However, for the iris data, no results for any *D* showed any significant difference, see table 2.

In addition, the FP rate increased as *D* increased. The paired sample t-tests were performed on the different pairs which were made in the same way as previous paired sample t-tests. The t-tests showed that $D = 1$ and $D = 5$ was statistically significant with 94.7% confidence. A two-sided t-test of the means produced a p-value of 5.2177%. Much larger p-values were produced when the t-tests were performed on the rest of pairs, ranging from 35.7729% to 98.7759%. These results also show that the FP rate increased as the detector sample size increased but that it stabilised to a certain point.

5.3.2 Analysis

The observed results were expected. When a detector sample size is one, no niching mechanism can happen. Since there is no chance for a selected detector to compete with other detectors to gain a fitness score, each detector will increase its fitness score by one as long as it matches a given antigen (when A=1). Thus, the generalist detector, which detects the largest number of non-self antigens during the fitness scoring procedure, will have the highest fitness score (assuming that each detector is selected with the same probability). Conversely, more specific detectors will gain much lower fitness scores in the same generation since they will detect much fewer non-self antigens. Thus, the generalist detectors will dominate in a detector population after a certain number of generations.

This kind of phenomenon resulted in rather poor results for the cancer and vote data when D = 1. However, the detector sample size did not affect average TP rates for the iris data at all. This is perhaps because the given problem of iris data is relatively easier and thus the minimum sample size is good enough to show a good detection rate. In other words, fairly general detectors can detect all existing non-self antigen patterns in the iris data set.

When the detector sample size is more than one, the selected sample detectors compete with each other. In our tests, this led the winner detectors from sampled detector groups to form niches, which match separate peaks of a fitness landscape. In the extreme case, when the detector sample size is the largest possible (the detector population size), every detector participates in a competition to detect a given antigen. This gives a chance for very specific detectors to increase their fitness scores because some specific non-self antigen patterns can only be detected by these kinds of detectors. Therefore, these specific detectors will have fitness scores that are large enough not to be excluded from the parent population through selection. In other words, both the general detectors and specific detectors have fair chances to win and thus they both will remain in the final detector population.

However, when a detector sample size is the largest possible, it can cause an overfitting problem. The specific non-self antigen pattern may not be representative of the data as a whole. So a detector evolved to match this exceptional antigen pattern might not truly distinguish between “self” and “non-self”, resulting in higher false-positive rates. This overfitting problem is clearly observed from our experiment results. For both data, cancer data and vote data, the FP rate increases as the detector sample size increases, see table 1.

5.3.3 Varying Antigen Sample Size

We next compared the results when the detector sample size was fixed but the antigen sample size changed. The last series of experiments were performed with D = 10 and various antigen sample sizes. As seen in table 3, no significant difference between TP’s and FP’s was evident, except for the case where the antigen sample size was the maximum.

5.3.4 Analysis

These results are also readily explainable. When the antigen sample size is small, even a potentially general detector does not have enough opportunity to detect a large number of antigens and thus both a general detector and a specific detector will be compared only for whether they can detect a given small number of antigens. Thus, the difference of fitness scores is not large. However, as the antigen sample size increases, the general detector starts to have enough chances to beat the specific detector by detecting a larger number of antigens. Thus, the general detectors have more chances to be selected as the parents for the next generation. So larger antigen sample sizes can also cause domination of general detectors during evolution.

5.3.5 Performance of Negative Selection Operator

Finally, we observed that the negative selection operator played an important role which helped to reduce the FP rate. When evolution terminated at the maximum generation and the detectors were tested on a training data set, no case showed any mistake, ie, FP was always 0% on the training data set. For the test set, the observed FP rate was up to about 10%.

As discussed before, the FP rate is mainly controlled by a detector sample size and an antigen sample size. Therefore, the rather higher FP rates resulted not because of inappropriate behaviours of the negative selection operator but because of the improper choices of detector sample sizes and antigen samples sizes. However, we have not investigated how the threshold size of negative selection operator will affect the TP and FP rate. Too small a threshold size might lead to prevent the generation of some general detectors because it will eliminate detectors matching very small number of self antigens. However, these self antigens can be noise. Similarly, too large a threshold size can make the AIS to generate the detectors which are so general that they detect too many self antigens. Thus, the effect of negtative selection threshold size should be investgated as the future work.

5.4 Ideal Detector and Antigen Sample Sizes

Since an ideal IDS should show a high TP rate and low FP rate, we analysed TP-FP rates to take into account these two rates together. As shown in table 1, for cancer data, these rates did not show significant differences for any case. For the vote data, it stabilised after a detector sample size reached 10. These results advise that the detector sample size does not have to be the largest one to get the most ideal result. Instead, we can set a detector sample size

that is not too small but is large enough to gain the good TP-FP rate. To be more precise, we suggest that the detector sample size should be set as the largest size which is affordable by given system resources. As future work, an adaptive sample size determined through evolution can also be investigated.

As long as the detector sample size is properly set, the antigen sample size is not critical. Since our experiment results show that the generality of detectors can be controlled by the detector sample size, the smallest antigen sample size (A = 1) is recommended. This is because the minimum antigen size saves computation time.

6 Conclusions

This paper has investigated the use of a static clonal selection algorithm with a negative selection operator as one component of the AIS for network intrusion detection. This component was especially developed for the purpose of building a misuse detector in a more efficient way. In order to adapt the available clonal selection algorithm for a network intrusion detection problem, three major modifications were made: i) new genotype and phenotype representations, ii) new matching function and fitness score function and iii) introduction of a negative selection operator.

Two series of experiments were performed by varying a detector sample size and an antigen sample size. These experiments were performed in order to investigate the effect of detector and antigen sample sizes on performance. The first series of experiments proved that both the TP rate and the FP rate increases as the detector sample size increases. The second series of experiments also showed that the significant differences of TP and FP rates were observed only for the case that the antigen sample size is the maximum. Furthermore, the negative selection operator embedded in the clonal selection algorithm performs well from the two sets of experiments.

As the result of these experiments, this paper suggests that the largest detector sample size should be selected from the sample size range affordable by given system resources. Moreover, the antigen sample size is not critical as long as the detector sample size is properly set and thus the smallest size, which is one, will be ideal to save computation time.

Bibliography

D'haeseleer, P, (1997), "A Distributed Approach to Anomaly Detection", *ACM Transactions on Information System Security*. http://www.cs.unm.edu/~patrik/

Dasgupta, D., (1998), "An Overview of Artificial Immune Systems and Their Applications", In Dasgupta, D. (editor). *Artificial Immune Systems and Their Applications*, Berlin: Springer-Verlag, pp.3-21.

Dougherty, J. et al, (1995) "Supervised and Unsupervised discretisation of Continuous Features", *the Proceeding of 12th Int. Conference on Machine Learning*, pp.194-202, 1995.

De Jong, K., et al, (1993) "Using Genetic Algorithms for Concept Learning", *Machine Learning*, Vol.13, No.2/3, pp.161-188.

Fayyad, U. M., and Irani, K. B., (1993) "Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning", *Proceeding of The Thirteenth International Joint Conference on Artificial Intelligence*, pp.1022-1027.

Forrest, S. et al, (1993), "Using Genetic Algorithms to Explore Pattern Recognition in the Immune System", *Evolutionary Computation*, 1(3), 191-211.

Forrest, S., et al, (1997), "Computer Immunology", *Communications of the ACM*, 40(10), 88-96.

Hart, E. and Ross, P., (1999), "An Immune System Approach to Scheduling in Changing Environments", GECCO'99, pp.1559-1566, 1999.

Hofmeyr, S. And Forrest, S., (2000), "Architecture for an Artificial Immune System", *Evolutionary Computation*, vol.7, No.1, pp.45-68.

Kim, J. and Bentley, P., (1999a), "The Human Immune System and Network Intrusion Detection", *7th European Conference on Intelligent Techniques and Soft Computing (EUFIT '99), Aachen, Germany*.

Kim, J. and Bentley, P., (1999b), "The Artificial Immune Model for Network Intrusion Detection, *7th European Conference on Intelligent Techniques and Soft Computing* (*EUFIT'99), Aachen, Germany*.

Kim, J. and Bentley, P., (2000), "Negative Selection within an Artificial Immune System for Network Intrusion Detection", *the 14th Annial Fall Symposium of the Korean Information Processing Society, Seoul, Korea*.

Mykerjee, B., et al, (1994), "Network Intrusion Detection", *IEEE Network*, Vol.8, No.3, pp.26-41.

Somayaji, A., et al, (1997), "Principles of a computer immune system", *Proceeding of New Security Paradigms Workshop, Langdale, Cumbria*, pp.75-82.

Smith, R. E., et al, (1993), "Searching for Diverse, Cooperative Populations with Genetic Algorithms", *Evolutionary Computation*, 1(2), 127-149

Tizard, I. R., (1995), *Immunology: Introduction*, 4th Ed, Saunders College Publishing.

Witten, I. H. and Frank, E. (2000) *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann Publishers.