

Chapter 8. Conclusion

8.1 Introduction

This thesis has focused on studying the application of combinations of different artificial immune algorithms to intrusion detection. It began with a wide-ranging literature study with two topics: intrusion detection systems and immune systems. The literature study provided insight into the design goals for an effective network-based IDS, and the human immune features that can usefully contribute to its development. Based on this understanding, a novel artificial immune model was proposed integrating three evolutionary stages. In order to demonstrate the benefits of an integrated artificial immune model for IDS, the limits of an individual immune algorithm were identified. Following the identification of these limits, the role of each different artificial immune algorithm within an integrated system was studied. Then, an integrated artificial immune system was developed combining negative selection, clonal selection and gene library evolution. The integrated system was thoroughly evaluated to determine whether it is able to provide adaptability. The results of this evaluation have shown that the system is able to adapt to continuously changing environments, dynamically learning the fluid patterns of ‘self’ and predicting new patterns of ‘non-self’.

A list of research goals that support the research hypothesis is given in the introduction to this thesis. This chapter reviews the main contributions of the thesis by revisiting this list. The review highlights the actual outcomes shown throughout the thesis and considers whether those outcomes verify the achievement of the research goals. Following the review of thesis contributions, suggestions are made for future work on improvement of the AIS that was developed for the thesis. The suggested future work is explained in terms of intrusion detection applications and new artificial immune modelling. Finally, the thesis concludes with description of the new understanding and advances achieved by this study.

8.2 Review of Thesis Contributions

- 1. Identify crucial components of human immune systems (HIS) that can contribute to the improvement of AIS for applications such as intrusion detection.**

In order to define a new artificial immune model for network intrusion detection, literature related with two fields, intrusion detection and immune systems, was extensively reviewed (Chapter 2). Based on this review, a set of general requirements for network-based IDS’s and three principal design goals satisfying these requirements were identified (Chapter 3).

The identified three design goals are *distributed*, *self-organising* and *lightweight*. The three design goals allowed this research to analyse the salient features of human immune systems that can contribute to the design of effective network-based IDS's. Analysis through literature study has shown that coordinated actions of several sophisticated mechanisms of the human immune system satisfy all the identified design goals. Consequently it was possible to identify nine particular sub-components within the human immune system that harness features beneficial to IDS's. These components and the properties provided by them are listed in table 8.1.

| HIS Components | Property Provided by these Components | Requirements of Network-Based IDS |
|---|---------------------------------------|---|
| Immune Network Unique Antibody Sets | Distributed Detection | Robustness Configurability Extendibility Scalability |
| Clonal Selection Negative Selection Gene Library Evolution | Self-Organising | Adaptability Global Analysis |
| Approximate-Binding Gene Expression Somatic Hypermutation Memory Cells | Lightweight | Efficiency |

Table 8.1 The crucial components of HIS, the properties they provide and the requirements of network-based IDS

To be more precise, an *immune network* and *unique antibody sets* can provide the distributed property of IDS's, *clonal selection*, *negative selection* and *gene library evolution* can allow IDS's to be self-organising, and *approximate binding*, *gene expression*, *somatic hypermutation* and *memory cells* can make IDS's lightweight.

2. Propose an integrated new artificial immune model

Understanding of which are the crucial components of human immune systems led to a proposal for a new artificial immune model (Chapter 3). The proposed model embeds the useful immune components into one system (figure 8.1). It consists of a primary IDS and several secondary IDS's. It combines three evolutionary stages. *Gene library evolution* simulates the first stage of evolution, in which knowledge is gained of currently existing antigens. This process is driven by *gene expression*. The *gene expression* process generates various *immature detectors* by applying various genetic mechanisms on detectors which are

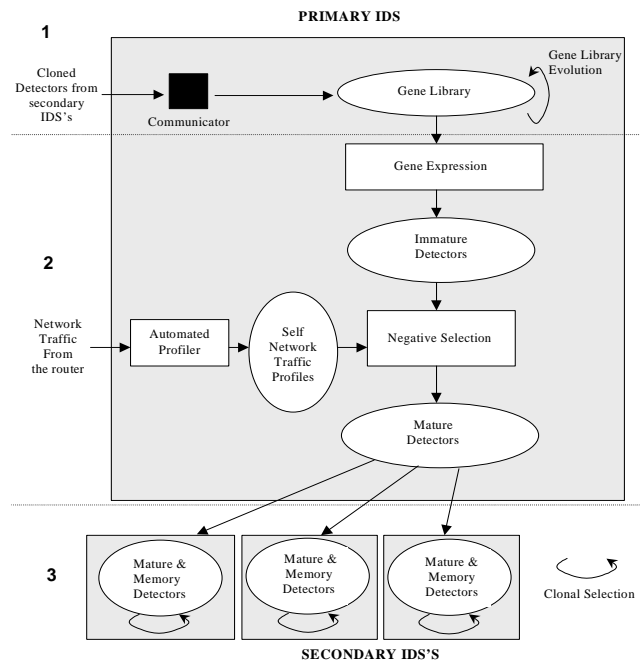


Figure 8.1 The Proposed Artificial Immune Model

randomly selected from the gene library. These mechanisms permit the artificial immune model to detect numerous intrusions using a smaller number of detectors, making it lightweight in a self-organising way. *Negative selection* forms the second stage of evolution, selecting mature detector sets by eliminating false immature detectors in a self-organising way. The transfer of unique detector sets to the secondary IDS's also occurs at this stage, making the model distributed. *Clonal selection* is the third stage of evolution, detecting various intrusions with a limited number of detector sets using approximate binding, and generating memory detectors. This generality and efficiency results in the model being lightweight. In addition, this process drives the *gene library evolution* in the primary IDS. These three processes are co-ordinated across a network to satisfy the three goals for designing effective IDS's: being distributed, self-organising and lightweight.

3. Identify limitations of the current AIS that is popularly used for intrusion detection

In order to demonstrate the benefits of a new integrated artificial immune model for IDS, the limits of individual artificial immune algorithms were identified (table 8.2). The artificial immune model proposed in the thesis was not the first attempt to develop an AIS for network intrusion detection. Various approaches to build an AIS have been attempted, mainly by implementing only a small subset of the human immune mechanisms. The omission of crucial components in order to simplify the development and application of the

AIS may detrimentally affect the effectiveness of an AIS. This potential problem has been shown by the experiments performed in Chapter 4.

| Limitation | Details |
|-----------------------|--|
| Scaling problem | When the negative selection algorithm is applied to a broader range of self, as in a realistic network environment, it requires generation of exceptionally large amounts of detectors and causes an unacceptably long computation time. |
| r-contiguous matching | Antigen-Antibody matching function is too simple to be applied to detect rather complicated and high-dimensional patterns in network traffic. |

Table 8.2 Limitations of the current AIS, the negative selection (NS) algorithm, for use in intrusion detection

The negative selection algorithm has been the most popular approach to applying a human immune mechanism to network intrusion detection (see Chapter 2). The feasibility of the negative selection algorithm in a real network environment was examined in Chapter 4. For the feasibility test, a broader range of self sets are defined, which can cover a more extensive range of network intrusions. In addition, two major changes were made in order to cope with these more complicated self sets, namely: the adoption of larger cardinality genotypes and the application of a matching function on phenotypes rather than genotypes. The results of experiments performed on this new self set showed a severe scaling problem for the negative selection algorithm. When the self definition widened, the string needed to encode a detector lengthened. As the result of the long length of detectors, the number of detectors required to gain an acceptable false negative error became huge, and thus required an unacceptably long computation time.

Another identified drawback of the negative selection algorithm was the adoption of the r-contiguous rule to check the match between a given detector and antigen. The negative selection algorithm requires an appropriate number of detectors in order to produce acceptable error and detection rates. The established formula that approximates the appropriate number of detectors is applicable only when the algorithm uses the r-contiguous matching function. However, the r-contiguous matching rule was too simple to be used for determining the match between complex and high-dimensional patterns. Since the r-contiguous bit matching only measured the contiguous bits of genotypes of the two given strings, it was hard to guarantee that it can detect correlations in complex self and non-self patterns.

4. Understand the role of each different artificial immune algorithm as a single component of an integrated system.

The drawbacks of the negative selection algorithm caused the AIS struggle to monitor the vast self set of a network despite its other important features. While the negative selection algorithm made the AIS an invaluable anomaly detector, it was infeasible to apply it to a real network environment. To be more precise, the negative selection stage should be restricted to tackle a more modest task that is closer to its role in the human immune system. In the human system, it simply filters the harmful antibodies rather than generating competent ones. Consequently, the experimental results obtained in Chapter 4 suggest the need to re-define the role of the negative selection stage within a network-based IDS, and to design a more applicable negative selection algorithm according to that new role.

Following the redefinition of the role of the negative selection stage, the second evolutionary stage, clonal selection, has been studied in Chapter 5. In particular, the clonal selection algorithm developed in Chapter 5 includes a negative selection operator. This new algorithm, called the static clonal selection algorithm (StatiCS) lets detectors evolve towards the non-self patterns hidden in the collected non-self data. This component was developed especially for the purpose of building a misuse detector in a more efficient way. Table 8.3 shows the roles of two significant immune algorithms: negative selection and the static clonal selection algorithm.

| AIS Algorithm | Role |
|-----------------------------------|-----------------------------|
| Negative Selection | Filtering invalid detectors |
| Static Clonal Selection Algorithm | Misuse detector |

Table 8.3 The Roles of different artificial immune algorithms

From the development of the static clonal selection algorithm and related experiments, the following three contributions are made.

- **A modified representation of detectors and a matching function for a static clonal selection are introduced. These no longer require the arbitrary choice of the matching threshold value, to which the detection rate is sensitive.**

The StatiCS is the advanced clonal selection algorithm introduced in [Smith *et al.*, 1993]. In order for the StatiCS to be applied to the network intrusion detection problem, three major modifications were made: i) new genotype and phenotype representations, ii) new matching and fitness score functions and iii) introduction of a negative selection operator. Although

these modifications are not entirely novel and they can be found from in the concept learning research field, their application to network intrusion detection within an AIS framework is arguably the first reported attempt. Consequently, a new phenotype representation (representing a conjunctive rule) removes the matching threshold parameter by allowing an “OR” operator in a genotype-phenotype mapping. The matching threshold parameter has been known to affect greatly the detection rate of the original negative selection algorithm, but the choice of parameter values has been made arbitrarily. In contrast, the modified phenotype representation can dynamically alter the matching scope of a given detector since it embeds an “OR” operator in the phenotype. Thus, the modified phenotype representation no longer requires the arbitrary choice of a parameter value that significantly affects the detection rate. In addition, this phenotype allows a detector to detect more than one specific antigen pattern and thus it still have lightweight feature originated from approximated binding. Furthermore, the detector phenotypes used in this work will have a larger degree of intelligibility. This is because they do not require a numerical threshold whose actual meaning is hard to be understood. When an IDS is designed to send non-self detections to a security officer to confirm and draw a reaction, the intelligibility of detector phenotype is one of the most significant IDS requirements to be satisfied.

- **A static clonal selection algorithm demonstrates efficient niche maintenance.**

Previous work [Forrest *et al.*, 1993; Smith *et al.*, 1993] has shown that two important factors influence the degree of antigen pattern recognition by the StatiCS employed in this thesis. These two factors, the detector and antigen sample sizes, are investigated in order to generate an appropriate balance between general and specific detectors for learning non-self antigen patterns. Since the AIS used in all previous work employed a binary detector and simple matching functions (such as Hamming distance or r-contiguous matching), a new investigation was made in order to understand whether the StatiCS with the modified detector representation and matching function still allow it to maintain an efficient niching strategy. Two series of experiments were performed by varying the detector and antigen sample sizes. The results of these experiments show a good non-self antigen detection rate with a relatively large detector sample size. Moreover, the antigen sample size is not critical as long as the detector sample size is set optimally. Thus the antigen sample size is set to one, the smallest possible value, in order to save computation time. These results also suggested that a more precise mixture of general and specific detectors causes the algorithm to follow the fitness landscape more closely as the detector sample size increases.

- **A static clonal selection algorithm with a negative selection operator demonstrates acceptable self-tolerance.**

The experimental results of the StatiCS also showed a low self antigen detection rate. In contrast with the results obtained for the negative selection algorithm, these results were gained in acceptable times. This work has shown that clonal selection provides effective non-self antigen detection and that the most appropriate use of negative selection in the AIS is as a filter for invalid detectors, not the generation of competent detectors.

5. Study new effects and limitations of an integrated model

| New Effects |
|---|
| <ul style="list-style-type: none"> • Incrementally learning globally converged normal behaviours by being exposed to only a small subset of self antigens at one time. • Replacing detectors effectively when the converged behaviours learned in an incremental way are suddenly altered due to legal self change. • Gaining high TP rates without increasing the amount of costimulation as the result of gene library evolution |
| Limitations |
| <ul style="list-style-type: none"> • Human involvement • Noisy input handling |

Table 8.4 New effects and limitations of an integrated model, DynamiCS

As the final stage of work, this research developed a dynamic clonal selection algorithm (DynamiCS) that controls the proliferation and extinction of detectors within IDS's. DynamiCS is an integrated model that combines negative selection, clonal selection and gene library evolution. Table 8.4 summarises the new effects and limitations of DynamiCS. These new effects and limitations are described in the following section.

- **DynamiCS learns normal behaviours by being exposed to only a small subset of self antigens at one time.**

The self-organising feature of DynamiCS was tested by simulating specific scenarios that commonly occur with IDS's. One such case is learning of normal behaviours by being exposed to only a small subset of self antigens at any one time. In order to equip DynamiCS with this property, additional human immune features were identified. They are central tolerisation, distributed tolerisation, costimulation, affinity maturation, life span and memory detectors. DynamiCS implemented these features by introducing three important

parameters: tolerisation period, activation threshold and life span. The experimental results showed that the DynamiCS was able to incrementally learn the globally converged distributions when only a small subset of antigens was given at each generation. The three parameters influence the non-self antigen detection (TP) and self-tolerance (FP) rates significantly. An analysis determined that TP and FP rates vary depending on the number of detector activations in total, and that this number was directed by values of the three parameters. This is because DynamiCS generated initial immature detectors only through negative selection, and as a result the degree of antigen detection did not vary greatly between detectors. In addition, it was observed that a larger tolerisation period caused the system to produce a smaller number of detectors to activate and resulted in lower TP and FP rates. In contrast, different values of activation threshold did not affect the number of detectors activating, but a larger activation threshold made mature detectors activate much less frequently, and ultimately triggered a smaller number of detector activations, also bringing about lower TP and FP rates. In addition, larger values for the life span of mature detectors forced mature detectors to live longer and experience more antigens. This led the system to provoke more detector activation, which produced higher TP and FP rates.

- **The extended DynamiCS replaces detectors reflecting previously observed normal behaviours but no longer representing current normal behaviours.**

DynamiCS was also tested to determine whether it can replace detectors effectively when the converged behaviours learned in an incremental way are suddenly altered due to legal self change. The experimental results showed that large tolerisation period values, that were sufficient to show perfect FP rates in previous experiments, no longer demonstrated perfect FP rates. This was because generated memory detectors had never been exposed to a certain antigen cluster and thus they could not have perfect self-tolerance. This reason led the further extension of DynamiCS, to handle generated memory detectors based on their detection results. The extended DynamiCS eliminated memory detectors when they showed poor self-tolerance to new antigens. The experimental results showed that deletion of memory detectors based on their self-antigen detection dramatically decreased high FP rates.

- **Gene library evolution using hypermutation reduces the amount of costimulation by fine-tuning deleted memory detectors.**

The elimination of memory detectors by the extended DynamiCS still required a larger amount of costimulation in order to lower FP rates. A large amount of costimulation (human intervention) can render the system useless for intrusion detection. In order to resolve this

problem, DynamiCS employed the use of hypermutation in DynamiCS to produce the effect of gene library evolution. This additional extension was designed to fine-tune generated memory detectors so that the system obtained higher TP rates without increasing the amount of co-stimulation. The gene library evolution was modelled by producing immature detectors via hypermutation on deleted memory detectors. Thus a “virtual gene library”, made from mutations of deleted memory detectors was maintained. The new extension was tested to determine whether it gains high TP rates without increasing the amount of costimulation as the result of gene library evolution. The test results proved that hypermutation led the progress of gene library evolution and thus produced immature detectors that are more tuned to cover existing non-self antigens. These were understood by analysing different *degrees* of TP rate drop when gene library evolution was and was not applied. With gene library evolution, DynamiCS showed a smaller TP rate drop when antigens were presented from the same antigen cluster. This is because the mutants of previously deleted memory detectors, having survived the negative selection stage, are likely to have some non-self antigen information without patterns matching self antigens.

- **Limitations of DynamiCS**

Although DynamiCS has shown several new effects, it also has limitations. Firstly, DynamiCS cannot achieve fully automated anomaly detection. The human involvement plays an important part in obtaining a satisfactorily low false positive error rate. The gene library evolution effect decreases the degree of human involvement. However, it does not eliminate the human involvement. This limitation is related to the self/non-self discrimination assumption which is applied to DynamiCS. Since DynamiCS regards antigens that are not self as intrusions, it is bound to make some mistakes by detecting antigens that are “not self” but not intrusions either. This limitation originates from the fundamental assumption that continues to be discussed actively by immunologists. Radically different models such as the danger model explain the human immune response without using the self-non-self discrimination concept [Matzinger, 1994].

Secondly, DynamiCS makes an assumption that there is no noise in antigen sets provided for negative selection. When an antigen set including both self and non-self antigens is transferred to the system, if non-self antigens are not detected by memory detectors and activated mature detectors, those non-self antigens are treated as self antigens. Thus, these non-self antigens become noise for the purposes of a negative selection. This immediately affects the system’s likelihood of providing false self information and thus results in the elimination of sound immature detectors. In other words, it causes the loss of some potential mature detectors that could have detected non-self antigen patterns. Therefore, the

lower non-self antigen detection rates caused by antigen noise greatly depend on how quickly memory detectors and activated mature detectors detect hidden non-self antigen patterns.

8.3 Future Work

The promising results of this thesis provide an important step forward in the development of artificial immune systems for intrusion detection. This certainly drives several items of important future research into developing more competent AIS's.

8.3.1 Evaluate on a Real Network Data Set

Although the integrated artificial model that is implemented in this thesis has shown good adaptability to continuously changing non-self and self patterns, the developed system has not been tested on real network traffic data. The first obvious step of future work would be the extensive assessment of the StatiCS and DynamiCS on a real network traffic data set. The StatiCS needs to be tested to see if it can be an efficient misuse detector, and DynamiCS also needs to be evaluated to see whether it still can be adaptive to real network traffic data. As in the evaluation of the negative selection algorithm in Chapter 3, a realistic range of self definition should be used for the test. Having experienced the exceptionally large amount of data that is usual for network traffic, scalability is clearly also a very important evaluation criterion. The current integrated model, DynamiCS, also employs several mechanisms that can allow it to be lightweight. A lightweight AIS does not impose a large overhead on a system or place a heavy burden on CPU and I/O. If an AIS is lightweight, it will certainly increase the scalability of the system. Hence, it would be important to analyse to what extent the lightweight property of the AIS can contribute to increase the scalability of the system. The evaluations of the StatiCS and DynamiCS performed in this thesis are concentrated on the proof of concept. More confident claims about the system's performance as an IDS would be gained after more thorough tests on real network data. For this purpose, DynamiCS is currently being used to solve a different real world problem. The promising research results obtained from this thesis motivated a new research project to apply DynamiCS. The Computer Immunology for Fraud Detection (CIFD) project at King's College London (KCL) currently uses DynamiCS for financial fraud detection. More thorough evaluation of DynamiCS will be performed as a part of this project.

8.3.2 Distributed Detection

The artificial immune model originally proposed in Chapter 3 is expected to detect intrusions in a distributed way. As discussed in Chapter 3, a distributed IDS can provide robustness, configurability, extendibility and scalability. These significant properties can be obtained by delegating the intrusion detection responsibilities to a number of distributed components. A

number of independent intrusion detection processes monitor only a small aspect of the overall system. They can operate concurrently and co-operate with each other.

Hofmeyr's LYSIS [Hofmeyr, 1999; Hofmeyr and Forrest, 2000] is the first AIS developed for application to IDS that operates under a distributed environment. Nevertheless, the definition of distributed detection introduced in [Hofmeyr, 1999; Hofmeyr and Forrest, 2000] is restricted. It assumes that LYSIS operates only under a broadcast LAN environment. A broadcast LAN environment transfers identical input network packets to all the local hosts in a domain. Although LYSIS has different sets of detectors at local hosts, they are exposed to exactly the same set of input network packets. This kind of environment is a very special case. With a switched Ethernet, for example, each host only can experience network packets transferred to it and thus network packets handled by each detector set are different from each other. Due to this rather special circumstance, LYSIS was able to achieve several novel features such as scalability originated from the absence of communication among different detector sets, and robustness. This implies that no AIS having truly distributed components has been developed for IDS. The potential advantages of a distributed IDS have only been discussed theoretically in the literature, including the artificial immune model proposed in this thesis and some other work [Dasgupta *et al.*, 1999b]. None of these proposed models has been tested in a real distributed environment. Further development of distributed AIS and its study would be an important research avenue to be pursued.

There are several issues to be studied in order to implement the distributed detection procedure within the proposed artificial immune model. Firstly, what sort of communication between the primary IDS and the secondary IDS's, and among the secondary IDS's, will be required to perform global analysis? Will this communication be simple enough not to impose an unacceptable extra burden on the network? Secondly, how many cloned detectors at each host will be transferred to other hosts, and which hosts will these cloned hosts be transferred to? Can we group hosts depending on their user groups, or usages, and decide that only particular groups of hosts receive cloned detectors? Thirdly, will these distributed mechanisms indeed show the desired features identified in Chapter 2, such as robustness, configurability, extendibility and scalability. These features were achieved in LYSIS since LYSIS does not require communication between detectors at each host. However, under a more common environment where each host is presented with different network packets, some form of communication between different hosts will be necessary. Then, will these useful features still be obtainable?

8.3.3 Other Human Immune System Analogy

It is quite natural to move this work forward by exploring other mechanisms of the human immune system. Firstly, some human immune mechanisms included in the originally proposed

artificial immune model in Chapter 3 are not fully studied yet. These are the gene expression process and gene library evolution using a separate gene library, which collects only useful genes and not detectors. As discussed in [Kim and Bentley, 1999a], the gene expression process together with adoption of a separate gene library is expected to contribute to making the AIS lightweight (see Appendix B). These processes still need to be developed and studied. This stage of the work might be further extended by more carefully examining how a gene library is developed in the human body. The development of the gene library is another important research issue that is actively studied by computer scientists [Kumar and Bentley, 2000]. A good example of attempting to combine this with an AIS is demonstrated by [Bradley *et al.*, 2000] for a fault detection system. A similar idea can be used for AIS for IDS.

Another interesting work might be the adoption of Jerne's immune network theory for a distributed DynamiCS. As discussed in the previous section, one important research issue for developing a distributed model of AIS is to control the number of cloned detectors and their death and proliferation. Jerne's immune network theory describes how this mechanism emerges in a self-organising way. This model might provide some new ideas for controlling the appropriate number of detectors across different hosts.

8.3.4 Hybridisation with Other Algorithms

The remarkable features of the human immune system have been appealing to computer scientists for developing a new intelligent algorithm. As described in Chapter 2, Related Work, a major group of the AIS's that have been developed so far are focused mainly on mimicking the actual mechanisms of the human immune system. Whilst it is possible to draw an analogy between the human immune system and a desired artificial system, the two systems cannot be fully identical. This implies that the comprehension of differences between the two systems would be as important as the extraction of analogies between them. Considering that the artificial immune system has a relatively short history compared to other artificial intelligent techniques, it may be advantageous to compose AIS with other mature intelligent techniques. In particular, this type of advantage can be notable when the differences between the human immune system and the artificial system cause the artificial immune algorithms to struggle with real world problems. Hybridisation with mature algorithms could provide a more powerful solution than the sole adoption of artificial immune algorithms.

For instance, the scalability of current artificial immune algorithms might be greatly increased when they adopt other data mining algorithms that have been proven to be successful at handling a massive amount of data. One of main research topics that have been widely studied in the data mining field is scalability. There are several algorithms available that scan a massive amount of data within a reasonable time and produce some summaries of collected data.

Association rule and frequent episode mining algorithms [Agawal *et al.*, 1993; Mannila *et al.*, 1997] are examples of these algorithms. These algorithms can be used to pre-process raw network traffic data and thus reduce the amount of data that has to be handled by the AIS. These algorithms can be also used as a feature constructor that constructs useful features from raw network data for IDS [Lee, 1999a]. Such an algorithm can be a good front-engine, defining a more optimised self set to be passed to the AIS.

The second example of hybridisation could be the combination of negative selection and other conventional anomaly detection algorithms. Negative selection is an important stage in the context of allowing the AIS to be an anomaly detector. However, it shows a severe scaling problem when required to generate a sufficient number of detectors. This problem often leads the AIS to produce much fewer number of detectors than is required and thus results in very low detection rates. Nevertheless, it has an important strength over other conventional anomaly detection algorithms. Conventional anomaly detection algorithms detect anomalies based on what looks significantly different from their norm definition. One big problem of conventional anomaly detection algorithms is that they have to define the boundary of self. The main task of the conventional anomaly detection algorithm is understanding normal behaviours from collected raw data. Many different techniques define the normal behaviours in various ways. However, it is very difficult to discern accurate definitions of normal behaviours. If these algorithms have less than accurate definitions of normal behaviours, they tend to produce high false positive error rates. Likewise, the negative selection algorithm and conventional anomaly detection algorithms have shown two distinct and severe problems: high false positive errors and low detection rates. Thus, the hybridisation of these two approaches might be a good solution for anomaly detection. For instance, conventional anomaly detection algorithms could first be used to define self (i.e. normal behaviours). From the provided data, those that are not categorised into the self definition could be selected as input data for the negative selection algorithm. Then, the negative selection algorithm could generate immature detectors by testing them only on selected input data. By doing so, the scalability of negative selection would be increased, since a much smaller amount of self data will be passed to the negative selection algorithm. For the monitoring stage, conventional anomaly detection algorithms would always detect potential anomalies first. These potential anomalies would then be re-checked by the detectors generated by the negative selection algorithm. This is because the detectors are generated from self data excluded from the self definition of anomaly detection algorithms. Conversely, anomaly detection by the detectors alone would result in the detection of self defined by the conventional anomaly detection algorithms. However, those detectors would be useful to detect false positive errors made by the anomaly detection algorithm. This kind of combination would be useful for generating an initial immature detector population from a large

self data set, since the generation of an immature detector population cannot initially benefit from clonal selection and gene library evolution.

Another important example can be the appropriate adoption of other evolutionary algorithms. As the proposed model adopts three important evolutionary stages, the cornerstone of the AIS is evolution. This raises the research question of how the AIS studied here is different from other evolutionary algorithms. This kind of study would have to be more focused on identifying the unique advantages and disadvantages of AIS's compared to other evolutionary algorithms. For instance, the usefulness of clonal selection and gene library evolution introduced in this thesis has been demonstrated by learning new antigens as they appear. The relation between learning and evolution is one of the important research issues studied in the evolutionary computation community [Nolfi and Floreano, 1999]. It would be fruitful to investigate whether previous research results on the relation between learning and evolution might provide a better understanding of the evolutionary stages within AIS's.

8.3.5 Different Applications

Although the artificial immune model proposed in the thesis limits its application to intrusion detection, it has been reported that there are many other applications that can obtain benefits from AIS (see Chapter 2). DynmiCS, an integrated AIS introduced in this thesis, can be used directly for some other anomaly detection oriented applications. These include virus detection, fault detection and fraud detection. Currently, DynamiCS is being applied to a financial fraud detection problem. Computational Immunology for Fraud Detection (CIFD) is a new collaborative project between King's College London (KCL), Anite Government Systems Ltd (AGSL) and Consignia, plc. CIFD investigates the adoption of DynamiCS for fraud detection in retail business. Many intelligent computation techniques have been effectively applied and numerous financial organisations reported their successful implementation [Treleaven and Goonatilake, 1995]. However, most accounts of successful fraud detection are in the financial sectors, where there is relatively good expertise on potential and experienced frauds [AAAI, 2002]. Therefore, diverse intelligent computation techniques used for these applications are focused mainly on automatically extracting hidden rules representing fraud from data gathered when various frauds occurred [AAAI, 2002]. However, unlike the financial sector, the retail sector rarely has sufficient knowledge about potential frauds or previously experienced fraud data. This requires the retail sector to employ a different fraud detection approach from other conventional methods. In order to develop a fraud detection system to meet such requirements, an anomaly detection based system is needed. DynamiCS will be further improved to meet this need and subsequently tested on a real world problem.

8.5 Conclusion

This thesis has shown that an artificial immune system is effective for intrusion detection when it combines separate artificial immune algorithms into one system. The following specific conclusions are drawn from throughout this thesis.

- The human immune system has crucial components that can contribute to the improvement of AIS for intrusion detection.
- An effective artificial immune system for network intrusion detection can be defined by integrating the three evolutionary stages: negative selection, clonal selection and gene library maintenance.
- The negative selection algorithm used in this thesis shows a severe scaling problem when applied in a real network environment.
- The clonal selection algorithm with a negative selection operator provides efficient niche maintenance and acceptable self-tolerance with a relatively large number of detector samples. Moreover, the antigen sample size is not critical as long as the detector sample size is optimally set, and thus the smallest possible antigen sample size, one detector, will be ideal in order to save computation time. This shows the potential of the clonal selection algorithm with a negative selection operator as an efficient misuse detector.
- Detector encoding using rule representation removes the need for a detector matching threshold.
- The most appropriate use of negative selection in the AIS is as a filter for invalid detectors, not the generation of competent detectors.
- A dynamic clonal selection algorithm that combines three evolutionary stages allows the AIS to be adaptable to dynamically changing antigen behaviours. It initially learns normal behaviours while encountering only a small subset of self antigens at any one time. Appropriate values of three parameters, tolerisation period, activation threshold and life span, provide satisfactory TP and FP rates.
- In addition, the dynamic clonal selection algorithm replaces memory detectors whenever previously observed normal behaviours no longer represent normal behaviour. This greatly reduces the false positive error rates.

- Gene library evolution using hypermutation reduces the amount of costimulation (human intervention) by fine-tuning existing memory detectors.

Based on the above conclusions, this thesis has argued that the integrated artificial immune model combining three evolutionary stages indeed shows the adaptability that is the desired property of intrusion detection systems. The overall artificial immune model for network intrusion detection presented in the thesis is not the first attempt to develop an AIS for network intrusion detection. The previous attempts to build an AIS have mainly implemented only a small subset of the overall human immune mechanisms [Dasgupta, 1998a]. However, as seen in other immunology literature [Paul, 1993; Tizard, 1995], the overall immune reaction is the carefully co-ordinated result of numerous components, such as cells, chemical signals and enzymes. Therefore, the omission of crucial components in order to facilitate the development and application of AIS detrimentally affected its performance. To resolve this problem, a new artificial immune model has been thoroughly studied, integrating three immune algorithms corresponding to three evolutionary stages. It was shown that each evolutionary stage plays its own significant role in order to make the AIS adaptable to dynamically changing self and non-self behaviours.

In conclusion, the remarkable mechanisms of the human immune systems that protect our body against diverse harmful antigens indeed provide great insights into developing an artificial immune system for intrusion detection. As the human immune system is able to self-organise via three evolutionary stages, an artificial immune model harnessing these three stages demonstrates adaptability to continuously changing environments, dynamically learning fluid patterns of ‘self’ and predicting new patterns of ‘non-self’.