

Contents

| | |
|--|-----------|
| CHAPTER 1. INTRODUCTION | 16 |
| 1.1 MOTIVATION | 16 |
| 1.2 THESIS HYPOTHESIS | 18 |
| 1.3 THESIS GOALS | 19 |
| 1.4 THESIS CONTRIBUTIONS | 19 |
| 1.5 THESIS STRUCTURE | 20 |
| CHAPTER 2. RELATED WORK..... | 23 |
| 2.1 INTRODUCTION | 23 |
| 2.2 INTRUSION DETECTION SYSTEMS | 23 |
| 2.2.1 Taxonomy of Intrusion Detection Systems | 24 |
| 2.2.2 Anomaly Detection Systems | 27 |
| 2.2.3 Network-Based Intrusion Detection Systems | 35 |
| 2.3 OVERVIEW OF HUMAN IMMUNE SYSTEMS..... | 39 |
| 2.4 ARTIFICIAL IMMUNE SYSTEMS | 42 |
| 2.4.1 Negative Selection Algorithm..... | 42 |
| 2.4.2 Clonal Selection Algorithm..... | 43 |
| 2.4.3 Gene Library Evolution | 44 |
| 2.4.4 Immune Network Theory..... | 46 |
| 2.4.5 Immune Memory | 48 |
| 2.4.6 Artificial Immune System Applications..... | 49 |
| 2.5 ARTIFICIAL IMMUNE SYSTEMS FOR COMPUTER SECURITY | 50 |
| 2.5.1 Virus Detection | 50 |
| 2.5.2 Intrusion Detection..... | 53 |
| 2.5.3 Fault Tolerant Software | 56 |
| 2.4 SUMMARY | 58 |
| CHAPTER 3. AN ARTIFICIAL IMMUNE MODEL FOR INTRUSION DETECTION | 60 |
| 3.1 INTRODUCTION | 60 |
| 3.2 REQUIREMENTS OF NETWORK-BASED IDS'S | 61 |
| 3.3 THE DESIGN GOALS OF NETWORK-BASED IDS'S..... | 62 |
| 3.3.1 Distributed | 62 |
| 3.3.2 Self-organisation | 63 |
| 3.3.4 Lightweight..... | 63 |
| 3.4 HUMAN IMMUNE SYSTEM FEATURES FOR NETWORK-BASED IDS'S..... | 63 |
| 3.4.1 Distributed Model | 64 |

| | |
|--|-----------|
| 3.4.2 Self-organisation | 64 |
| 3.4.3 Lightweight | 65 |
| 3.5 ARTIFICIAL IMMUNE MODEL FOR NETWORK INTRUSION DETECTION..... | 66 |
| 3.5.1 Overview | 66 |
| 3.5.2 Primary IDS | 68 |
| 3.5.3 Secondary IDS | 70 |
| 3.5.4 Summary of Artificial Immune Model..... | 71 |
| 3.6 THESIS SCOPE | 71 |
| 3.7 SUMMARY | 72 |
| CHAPTER 4. NEGATIVE SELECTION ALGORITHM..... | 73 |
| 4.1 INTRODUCTION | 73 |
| 4.2 RELATED WORK..... | 74 |
| 4.2.1 Negative Selection of The Human Immune System | 74 |
| 4.2.2 The Negative Selection Algorithm | 74 |
| 4.3 ALGORITHM OVERVIEW | 75 |
| 4.4 NETWORK TRAFFIC DATA VS NETWORK INTRUSION SIGNATURE..... | 77 |
| 4.5 EXPERIMENT DESIGN | 79 |
| 4.5.1 Objective | 79 |
| 4.5.2 Data and Parameter Setting | 80 |
| 4.6 EXPERIMENT RESULT..... | 82 |
| 4.7 ANALYSIS | 83 |
| 4.8 SUMMARY..... | 86 |
| CHAPTER 5. STATIC CLONAL SELECTION ALGORITHM | 87 |
| 5.1 INTRODUCTION | 87 |
| 5.2 RELATED WORK | 87 |
| 5.2.1 Clonal Selection of the Human Immune System | 87 |
| 5.2.2 Clonal Selection Algorithms | 88 |
| 5.3 ALGORITHM OVERVIEW | 88 |
| 5.3.1 Providing Self and Non-Self Antigens..... | 89 |
| 5.3.2 Discretiser | 89 |
| 5.3.3 Genotypes and Phenotypes | 90 |
| 5.3.4 The Matching Function..... | 91 |
| 5.3.5 Fitness Scoring..... | 92 |
| 5.3.6 Reproduction and a Negative Selection Operator | 92 |
| 5.3.7 Genetic Operators | 93 |
| 5.4 EXPERIMENT DESIGN..... | 94 |
| 5.4.1 Objective..... | 94 |
| 5.4.2 Data and Parameter Setting..... | 95 |

| | |
|--|------------|
| 5.5 EXPERIMENTAL RESULTS 1:NON-SELF AND SELF ANTIGEN DETECTION RATES | 96 |
| 5.5.1 Varying a Detector Sample Size | 96 |
| 5.5.2 Analysis | 98 |
| 5.5.3 Varying a Antigen Sample Size | 99 |
| 5.5.4 Analysis | 99 |
| 5.5.5 Ideal Detector Sample Size and Antigen Sample Size | 99 |
| 5.6 EXPERIMENTAL RESULTS 2: THE DEGREE OF GENERALITY IN A DETECTOR POPULATION | 100 |
| 5.7 EXPERIMENTAL RESULTS 3: THE PERFORMANCE OF THE NEGATIVE SELECTION OPERATOR | 103 |
| 5.8 SUMMARY | 104 |
| CHAPTER 6. DYNAMIC CLONAL SELECTION ALGORITHM..... | 106 |
| 6.1 INTRODUCTION | 106 |
| 6.2 RELATED WORK | 107 |
| 6.2.1 Dynamic Non-Self Antigen Detection by Human Immune systems..... | 107 |
| 6.2.2 Dynamic Anomaly Detection by AIS | 109 |
| 6.3 DYNAMIC CLONAL SELECTION (DYNAMICS) ALGORITHM | 111 |
| 6.3.1 DynamiCS Overview | 113 |
| 6.3.2 Immature Detectors..... | 116 |
| 6.3.3 Mature Detectors..... | 117 |
| 6.3.4 Memory Detectors | 119 |
| 6.3.5 Controlling Detector Birth and Death | 119 |
| 6.3.6 Self and Non-Self Antigen Detection..... | 121 |
| 6.4 DYNAMIC CLONAL SELECTION EXPERIMENTS..... | 122 |
| 6.4.1 Objective..... | 122 |
| 6.4.2 Data and Parameter Setting | 122 |
| 6.4.3 Experiment Design..... | 124 |
| 6.5 EXPERIMENT RESULTS 1: EXAMINATION OF COMPLETE ANTIGEN DATA..... | 125 |
| 6.5.1 Effect of the Tolerisation Period | 125 |
| 6.5.2 Effect of Activation Threshold..... | 131 |
| 6.5.3 Effect of Life Span | 134 |
| 6.5.4 Analysis | 136 |
| 6.6 EXPERIMENT RESULTS 2: EXAMINING ONLY ANTIGEN SUBSETS | 137 |
| 6.6.1 Varying the Generation Numbers to Provide Antigens from a Same Cluster | 137 |
| 6.7 SUMMARY | 140 |
| CHAPTER 7. EXTENDED DYNAMIC CLONAL SELECTION ALGORITHM..... | 142 |
| 7.1 INTRODUCTION | 142 |
| 7.2 DYNAMICS REVISITED | 143 |
| 7.3 RELATED WORK: HANDLING MEMORY DETECTORS | 146 |
| 7.3.1 Human Immune Memory | 146 |

| | |
|---|------------|
| 7.3.2 Artificial Immune Memory | 148 |
| 7.4 EXTENDED DYNAMICS: KILLING MEMORY DETECTORS..... | 150 |
| 7.4.1 Algorithm Description | 150 |
| 7.4.2 Experiment Results | 151 |
| 7.5 RELATED WORK: GENE LIBRARY EVOLUTION USING HYPERMUTATION | 154 |
| 7.5.1 Gene Library Evolution by Human Immune Systems | 155 |
| 7.5.2 Gene Library Evolution by Artificial Immune Systems..... | 157 |
| 7.6 EXTENDED DYNAMICS: SIMULATING GENE LIBRARY EVOLUTION USING HYPERMUTATION..... | 159 |
| 7.6.1 Algorithm Description | 159 |
| 7.6.2 Experimental Results | 161 |
| 7.6.3 Experimental Analysis | 165 |
| 7.7 DISCUSSION OF DYNAMICS | 167 |
| 7.8 SUMMARY | 168 |
| CHAPTER 8. CONCLUSION | 170 |
| 8.1 INTRODUCTION | 170 |
| 8.2 REVIEW OF THESIS CONTRIBUTIONS..... | 170 |
| 8.3 FUTURE WORK | 179 |
| 8.3.1 Evaluate on a Real Network Data Set | 179 |
| 8.3.2 Distributed Detection | 179 |
| 8.3.3 Other Human Immune System Analogy | 180 |
| 8.3.4 Hybridisation with Other Algorithms | 181 |
| 8.3.5 Different Applications..... | 183 |
| 8.5 CONCLUSION | 184 |
| REFERENCES..... | 186 |
| APPENDIX A. DETAILED DESCRIPTION OF THE HUMAN IMMUNE SYSTEM..... | 201 |
| A.1. INTRODUCTION | 201 |
| A.1.1 Specific Recognition | 202 |
| A.1.2 Genetic Structure of Lymphocytes..... | 203 |
| A.1.3 Development | 205 |
| A.1.4 Activation..... | 208 |
| A.1.5 Evolution | 209 |
| A.1.6 Self Tolerance | 210 |
| A.1.7 Distributed Detection | 213 |
| A.1.8 Summary of Human Immune Systems..... | 214 |
| APPENDIX. B. DISCUSSION OF THE ARTIFICIAL IMMUNE MODEL..... | 217 |
| APPENDIX. C. THE FIELDS OF NETWORK TRAFFIC SELF-PROFILES | 219 |

| | |
|---|------------|
| APPENDIX. D. SETTING ACTIVATION THRESHOLD AND LIFE SPAN VALUES IN DYNAMICS | 220 |
| APPENDIX. E. PUBLICATION | 222 |

List of Figures

| | |
|--|-----|
| Figure 2.1. Taxonomy of IDS | 25 |
| Figure 2.2 Development of B-cells and T-cells (left). Clonal selection (right) | 40 |
| Figure 2.3. Jerne's Immune Network, [Timmis, 2001] | 46 |
| Figure 3.1 Physical Architecture of the Artificial Immune Model | 67 |
| Figure 3.2. Conceptual Architecture of the Artificial Immune Model..... | 67 |
| Figure 4.1 Gene Expression Process..... | 74 |
| Figure 4.2 Detector Set Generation of a Negative Selection Algorithm [Forrest <i>et al.</i> , 1994]..... | 75 |
| Figure 4.3 Non-Self Detection by a Detector Set | 75 |
| Figure 4.4 A Detector Phenotype and a Self Phenotype..... | 76 |
| Figure 5.1 An Overview of the AIS | 89 |
| Figure 5.2 Detector Genotype and Phenotype | 90 |
| Figure 5.3 Reproduction and Negative Selection..... | 93 |
| Figure 5.4 Degrees of detector generality when a detector sample size changes and an antigen sample size is fixed as one. | 101 |
| Figure 6.1 A Life of a Detector, [Hofmeyr, 1999]..... | 110 |
| Figure 6.2 Pseudo Code of the Dynamic Clonal Selection Algorithm..... | 112 |
| Figure 6.3 Immature Detector Generation | 113 |
| Figure 6.4 Mature Detector Generation at generation T | 114 |
| Figure 6.5 Mature Detector Activation | 115 |
| Figure 6.6 Monitor using Memory Detectors | 116 |
| Figure 6.7 TP and FP rates when T varies and $A = 100$, $L = 10$, $N = 1$ | 126 |
| Figure 6.8 TP and FP rates when T varies and $A = 5$, $L = 10$, $N = 1$ | 127 |
| Figure 6.9 TP rates between generation = 800 and generation = 899 when $T = 20$ | 130 |
| Figure 6.10 TP and FP rates when A varies with $T = 5$, $L = 100$, $N = 1$ (1)..... | 132 |
| Figure 6.11 TP and FP rates when A varies with $T = 5$, $L = 100$, $N = 1$ (2)..... | 133 |
| Figure 6.12 TP and FP rates when L varies and $T = 5$, $A = 150$, $N = 1$ | 135 |
| Figure 6.13 TP and FP rates when N varies and $T = 30$, $A = 100$, $L = 10$ | 138 |
| Figure 7.1. TP and FP rates when A varies and $T = 30$, $L = 10$, $N = 30$ | 144 |
| Figure 7.2 TP and FP rates when A varies and $T = 30$, $L = 10$, $N = 30$ without memory detectors | 145 |
| Figure 7.3 TP and FP rates when A varies and $T = 30$, $L = 10$, $N = 30$ after killing memory detectors..... | 152 |
| Figure 7.4 Gene Library Evolution by the Baldwin Effect | 157 |
| Figure 7.5 Gene Library Evolution Effect via Antibody Evolution | 157 |
| Figure 7.6 Gene Library Evolution in the extended DynamiCS | 159 |
| Figure 7.7 Modified Pseudo Code for the Extended DynamiCS | 160 |
| Figure 7.8 TP and FP rates when A varies and $T = 30$, $L = 10$, $N = 30$ with mutation rate = 0.1..... | 162 |
| Figure 7.9 TP and FP rates when A varies and $T = 30$, $L = 10$, $N = 30$ with mutation rate = 0.2..... | 163 |
| Figure 8.1 The Proposed Artificial Immune Model..... | 172 |
| Figure A.1 Antigen Recognition by a lymphocyte | 202 |
| Figure A.2 B-cell receptor structure | 204 |

| | |
|---|-----|
| Figure A.3 B-Cell Receptor Genetic Organisation, [Opera, 1999]. The gene fragments randomly selected from gene library are rearranged. The rearranged gene fragments are joined by the transcription process. The splicing process joins the constant region, C, to VDJ and produces mRNA. | 205 |
| Figure A.4 B-cell development..... | 206 |
| Figure A.5 B-Cell Activation..... | 208 |
| Figure A.6 B-cell Clonal Selection..... | 210 |
| Figure A.7 Jerne's Immune Network, [Timmis, 2001]..... | 212 |

List of Tables

| | |
|---|-----|
| Table 4.1 Self Profiles | 79 |
| Table 4.2 Number of required detectors, N_r and number of trials to generate required number of detectors, N_{r_0} when false negative error, P_f , and the threshold, r , of r -contiguous matching function are given. These numbers are calculated when a self string length, $l = 33$, an alphabet cardinality, $m = 10$ and the number of self strings, $N_s = 192$ | 81 |
| Table 4.3 Time is an avarage time of single detector generation and Trial is an average trial number to generate a single detector. The average values are followed by the standard deviations in parentheses. | 82 |
| Table 5.1 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. | 97 |
| Table 5.2 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. | 97 |
| Table 5.3 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. | 99 |
| Table 6.1 Detector Birth and Death in the Dynamic Clonal Selection Algorithm..... | 120 |
| Table 6.2 The features of Cancer Data | 122 |
| Table 6.3 Parameter values used for the experiments performed in 6.5.1 Effect of the Tolerisation Period | 125 |
| Table 6.4 Proportion of Three Different Types of Detector when T varies and $A = 100$, $L = 10$, $N = 1$. The values in parentheses are variances..... | 129 |
| Table 6.5 Proportion of Three Different Types of Detectors when T varies and $A = 5$, $L = 10$, $N = 1$. The values in parentheses are variances. | 129 |
| Table 6.6 Parameter values used for the experiments performed in 6.5.2 Effect of the Activation Threshold | 131 |
| Table 6.7 Proportion of Three Different Types of Detector when A varies and $T = 5$, $L = 10$, $N = 1$ The values in parentheses are variances. | 134 |
| Table 6.8 Parameter values used for the experiments performed in 6.5.3 Effect of the Life Span | 134 |
| Table 6.9 Proportion of Three Different Types of Detector when L varies and $T = 5$, $A = 150$, $N = 1$. The values in parentheses are variances. | 136 |
| Table 6.10 Parameter values used for experiments performed in 6.6.1 Varying the Generation Numbers to Provide Antigens from a Same Cluster..... | 139 |
| Table 7.1 Parameter values used for DynamiCS experiments | 143 |
| Table 7.2 Average numbers of surviving, generated and deleted memory detectors during 2000 generations, and average number of memory detector costimulations per generation for the DynamiCS and the Extended DynamiCS. The values in parentheses are variances..... | 153 |

| | |
|--|-----|
| Table 7.3 Average numbers of surviving, generated and deleted memory detectors during 2000 generations, and average number of memory detector costimulations per generation for the extended DynamiCS with mutation rate = 0.1. The mean values are followed by the variances in parentheses. | 164 |
| Table 7.4 Average numbers of surviving, generated and deleted memory detectors during 2000 generations, and the average number of memory detector costimulations per generation for the extended DynamiCS with mutation rate = 0.2. The values in parentheses are variances. | 164 |
| Table 8.1 The crucial components of HIS, the properties they provide and the requirements of network-based IDS..... | 171 |
| Table 8.2 Limitations of the current AIS, the negative selection (NS) algorithm, for use in intrusion detection..... | 173 |
| Table 8.3 The Roles of different artificial immune algorithms..... | 174 |
| Table 8.4 New effects and limitations of an integrated model, DynamiCS | 176 |