How AIS Addresses Adaptability in IDS

Anazida Zainal, Mohd Aizaini Maarof and Siti Mariyam Shamsuddin
Faculty of Computer Science and Information Systems,
Universiti Teknologi Malaysia, 81310 Skudai, Johor
anazida, maarofma and mariyam@fsksm.utm.my

Abstract

Immune system is essential to human body as it protects our bodies from pathogens in an orchestrated manner. The attractive feature of this system is its ability to detect and fight against various kinds of pathogens. The adaptive process to match the shape of epitopes and gene library evolution manifest that the system is adaptive. Since body protection against pathogens is analogous to network protection against intruders, this has led many researchers to embark on the Intrusion Detection research deploying the Immune System approach. The discussion in this paper focuses on the aspects of adaptability of Immune System and how it is implemented in IDS. Three major works representing three major evolutionary processes in human immune system were chosen to be reviewed

1. Introduction

Various approaches have been taken towards ensuring the security of a system. Cryptographic approach and firewall are preventive method while intrusion detection system (IDS) complements this preventive mechanism. The main goal of intrusion detection is to detect unauthorized use, misuse and abuse of computers by both system insiders and external intruders [1][2]. Thus, a great challenge in computer security is to determine the difference between normal and potentially harmful activity [2]. Most of the intrusion detection systems either rely on the known attacks or the systems are exposed to a static set of normal network traffic during training. The second approach gives a high false alarm when the normal traffic pattern changes.

Researchers have used various methods to detect these potentially harmful activities. They span from expert systems, statistical approach [3], SVM [4] and artificial intelligence techniques such as Neural Network [5], Self Organizing Map [6] and Genetic Algorithm. The newly emerged technique called Artificial Immune System (AIS) is aggressively being researched in this area.

In the following sections, we briefly discuss the domain of IDS, AIS and its basic algorithms. Researches pertaining to the deployment of AIS to address the adaptability in IDS are also discussed.

2. Background

This section will give background knowledge both on IDS and AIS.

2.1 Intrusion Detection Systems

The idea of Intrusion Detection System was first conceived by Anderson [7] while working on the improvement of auditing facilities and surveillance abilities of computer systems. Following the idea thrown by Anderson, Denning [8] has proposed a generic framework of an Intrusion Detection System.
Generally IDS are categorized into 2 groups based on their monitoring scope and detection techniques. They are host-based and network-based IDS.

i) Host-based

Early IDS deployed host-based IDS where it monitors a single host and normally uses audit trails of a host operating system. Its either exists as a daemon process or it can be a separate IDS. According to [1], host-based IDS's can be referred to as stand-alone intrusion detection systems because their monitoring scope id restricted to only a single host.

ii) Network-based

Network-based IDS monitors any number of hosts on a network by scrutinizing the audit trails of multiple hosts [9]. Since attempted attack across the network cannot be addresses by this approach, it is necessary for an IDS to monitor multiple events generated on several hosts to integrate sufficient information. Thus, the use of network traffic information is more effective [1], 3 types of its architectural implementations are monolithic, hierarchical and co-operative.

There are two types of detection techniques, misuse and anomaly detections.

i) Misuse Detection

This approach defines suspicious misuse signatures based on system vulnerabilities, security policy and known attacks. Intrusion signatures are either manually encoded or automatically learned through data mining. Signature recognition techniques have a limitation in which they cannot detect novel intrusions whose signatures are unknown [3]. Its mechanism lies in the comparison of events in audit trail with the list of signatures. On the other hand, its advantage is that it has almost nil false positive.

ii) Anomaly Detection

Anomaly detection techniques capture both known intrusions and unknown intrusions if intrusions demonstrate a significant deviation from norm profile [3]. The cycle of anomaly detection begins with the establishment of a normal profile or normalcy model. Usually, the system will undergo training phase where normal data are exposed to the system. Once the normal profile is obtained, the training is considered complete. Thus, any significant deviations from this established normal profile can be considered anomalous. The advantage of this approach is that it has the ability to detect new intrusion (have not encountered before). But the drawback to this approach is high rate of false positive.

The trend shows that current IDS fuses both detection techniques as to get the strengths from them.

2.2 Adaptive IDS

Adaptability means the ability to respond to the changing environment. Many intrusion detection systems have been constructed by manual and ad hoc means. These systems have been designed and implemented based on the system builders' knowledge of a computer system and their understanding of known intrusions. As a result, the effectiveness and adaptability of the intrusion detection systems are limited in the face of new computing environments or newly attack methods [10]. There is a need to design and develop an IDS that is adaptive and efficient as to lessen the percentage of false positive as this may result to the downtime of a system for further diagnosis of a problem or false attack.

Also, there is an argument that the construction of the normal model and the detection operation which are usually carried out separately is no longer suitable due to the nature of the system itself. It is evolving, therefore, the model should be constructed periodically in order to provide a way of adaptation to the new environment [11].

In a real network environment, normal network traffic pattern often changes, thus making it difficult to observe a complete set of normal traffic data. There is a need to constantly updating the detectors or classifiers that can reflect the current pattern of normal traffic data [1]. Besides, the configuration of network often changes. Therefore, there is a need to design and construct an IDS which is capable of learning and adapting to the changes that occur. This will definitely contribute to an effective IDS.

Among the works related to the Adaptive IDS are in the area of Data Mining [10] and Machine Learning like AIS.

2.3 Artificial Immune Systems

Immune System's main function is to protect our bodies against constant attack of external microorganisms. It specifically recognizes and selectively eliminates foreign invaders by a process known as the immune response [12]. The bodies identify the invaders using two interrelate systems : innate and adaptive immune systems [13]. The former has the ability to recognize certain microbes and immediately destroy them. Whereas the adaptive immune system uses somatically generated antigen receptors generated by random process by concatenating gene segments. Each cell uses available segments differently to make a unique receptor, allowing the cells to collectively recognize malicious organisms confronted during a lifetime [13].

These attractive features of human immune system have led the researches to deploy the strengths from human immune system into wide range of application domains such as document classification, robotics, fraud detection, character recognition and network and host-based intrusion detection. These AISs have met with some success and in many cases have rivaled or bettered existing statistical and machine learning techniques [2]. Mimicking the immune system in protecting our bodies and fighting against invading pathogens are analogous to defending the system and network against attack in IDS. From the literature survey, most IDS works that deploy AIS fall under two different approaches, which are negative selection and Jerne's idiotypic network theory [2]. Most of the reported works use the former approach.

3. Immune Inspired Algorithms

In the following sections, 3 major evolution processes of human immune system and its corresponding algorithms in AIS will be discussed. They are; Negative Selection, Clonal Selection and Gene Library Evolution.

3.1 Negative Selection

The purpose of negative selection process performed by human immune system is to eliminate the immature detectors which bind to self cells. This training is taken place in the thymus. All the B-cells will be screened out to eliminate the detectors that mistakenly detect self cell as an invader. The detectors which pass the test, are released to roam the body and fight against invading pathogens. This process is similar to anomaly detection process in IDS. Below is the outline of the algorithm :

```
1.  Define self (normal traffic)
2.  Generate detectors
3.  Perform training
        For each detector, match the against self
            if they match (complement)
                eliminate the detectors
        else
            the detector becomes mature
4.  The mature detectors will monitor the occurrence of
    anomaly.
```

Few applications on virus detection [14], and anomaly intrusion detection [15] [1] were published using negative selection method.

3.2 Clonal Selection

The matured B-cells produced are not necessarily useful in detecting antigens because negative selection only checks and eliminates B-cells that have pattern similar to self (similar in this sense is its complement). To exclude these detectors, antigens will impose pressure on selected detectors. The detector with the highest affinity binding will proliferate and undergo somatic hypermutation and receptor editing process to better match the antigen shape (paratopes).

Clonal Selection algorithm as was first proposed by Castro and Zuben [17] and was referred as CSA. Later, they have improved the algorithm and it is called CLONALG [18]. Below is the CLONALG algorithm :

1. *Randomly generate an initial population of antibodies Ab. This is composed of two subsets Ab_m (memory population) and Ab_r (reservoir population)*
2. *Create a set of antigenic patterns Ag.*
3. *Select an antigen Ag_i from the population Ag.*
4. *For every member of the population Ab calculate its affinity to the antigen Ag_i using some affinity function. (e.g. Hamming Distance)*
5. *Select the n highest affinity antibodies and generate a number of clones for each antibody in proportion to their affinity, placing the clones in a new population C_i.*
6. *Mutate the clone population C_i to a degree inversely proportional to their affinity to produce a mature population $C_{i\_}$.*
7. *Re-apply the affinity function to each member of the population $C_{i\_}$ and select the highest score as candidate memory cell. If its affinity is greater than the current memory cell Ab_{mi}, then the candidate becomes the new memory cell.*
8. *Remove those antibodies with low affinity in the population Ab_r and replace them with new randomly generated members.*
9. *Repeat steps 3-8 until all antigens have been presented. This represents one generation of the algorithm.*

The Clonal Selection itself can be considered adaptive since the process will clone and evolve the best detector to match the given antigen population.

3.3 Gene Library Evolution

Learning in human immune system is when the Clonal selection process takes place as a response to the changing antigens. Somatic hypermutation is when process where a portion of genes that are randoml;y selected from antibody clone mutates. These mutants will be sent to parts of human body to compete with existing antibodies to identify antigens. And the highest affinity antibody will be chosen for cloning. According to Sompayrac [19], clonal selection with hypermutation is essential for the human immune system to permanently learn newly appearing antigens. Several works [20][21], aimed at finding the diversity required of a gene library in human immune system and the role of gene library evolution. It is found that antibody evolution gets slower and evolves to cover more random antigen niches when the pathogen size (exposed to antibodies) gets smaller. In this case, the immune system does not let the gene library evolve towards existing antigen specific niches. Instead, it evolves toward covering a coarse-grained antigen space. It can be concluded that gene library diversity is not maintained for recognition of specific pathogen, but it evolved to cover a coarse-grain encoding regions of pathogen population that the species encountered and learning is obtained through hypermutation that leads the immune system to fine-tune its detection of the existing antigen.

According to Kim [1] there are 2 methods currently deployed in order to evolve their gene libraries :

i) Gene Library Evolution by the Baldwin effect.

   Initially builds a gene library consisting of previously known antibody genes. The evolution of diversity later is obtained via gene expression and learning using hypermutation.

ii) Gene Library Evolution Effect via Antibody Evolution.

   Treats existing antibody population as a gene library and concentrates on antibody population evolution using learning through hypermutation.

4. AIS and Adaptability

All natural systems have the capability of coping with continuous changes in environment as to maintain life. Their survival capability includes adaptability, interactivity, self maintenance and diversity [22]. The requirement to be adaptive is the ability to learn. Human Immune System has the adaptive capability. For example, in the clonal selection and expansion process, the cells with mutated receptors of high affinity with the eliciting antigens are selected for survival. This illustrates that immune system learns to deal with a certain pathogenic agent by altering concentration and molecular structure of individual cells and molecules successful in combating specific disease [22].

Another phenomena of human immune systems that demonstrates the property of adaptability is the concept of network metadynamics, insertion of new cells and molecules into the network and removal of non-stimulated cells.

Therefore, the properties of human immune system like learning and adapting have led the development of the above algorithms.

The following sections will discuss the applications of the above algorithms in adaptive IDS.

4.1 The work of Adaptive IDS at University of New Mexico (15)

Hofmeyr's approach to tackle the issue of continuously changing environment was by extending the Negative Selection Algorithm. In contrast to other AIS work that produce detectors by monitoring static antigen set, his extended AIS created new detectors every day after the system experienced new network traffic which had not been presented before. The mechanism deployed would continuously compare the detectors to new antigens and the matching results of new antigen determine whether they have to be replaced by new detectors or not. The limited number of detectors population was refreshed to adapt to the new antigens set. The core feature of his work is the control of the life cycle of a detector according to its antigen matching results. The figure below illustrates the life of detector in his work. However the 2 day time frame was set to address the adaptability and to eliminate useless detectors and replenish the population with newly generated detectors. It is desirable to have the system that can detect any changes in the normal traffic and learn when necessary instead of continuously relearn in a fixed time period. Unfortunately, his system does not cater this need.

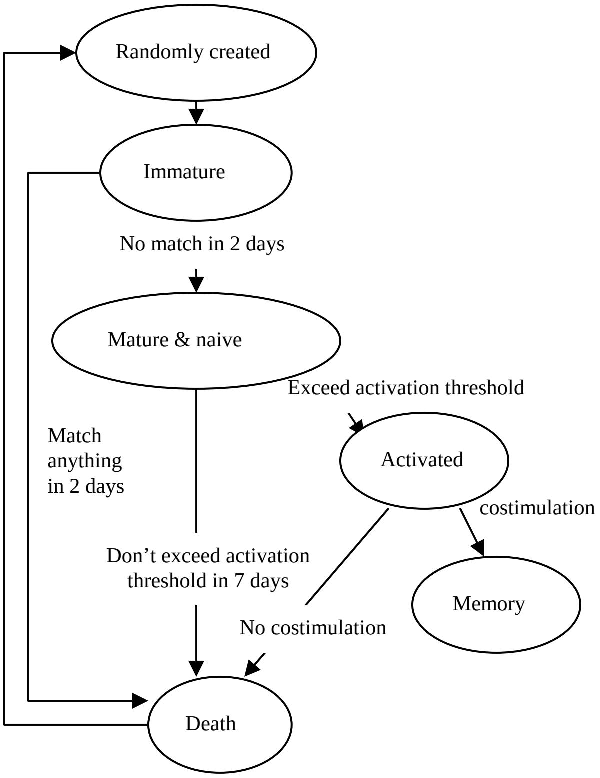

Figure 1 : The Life of a Detector [15]

In conclusion, his work addresses the aspect of adaptability via coordinated dynamics of three different detector populations : immature, mature and memory detector population [1].

4.2 The work of Adapative IDS at University College London [1]

Following the work of Hofmeyr, Kim [1] addresses the adaptability in intrusion detection from the Clonal Selection angle. Her modified version of Clonal Selection embeds Negative Selection algorithm in it and is called Dynamic Clonal Selection (DynamiCS). Like Hofmeyr, she also concentrated on the 3 fundamental elements that contribute to adaptability, which are immature detectors, mature detectors and memory detectors and the dynamics of the population. Her DynamiCS is summarized below :

Stage 1 : Initialization of detectors
Generation of immature detectors set.
    If immature detector binds to self, delete

```
Until # of immature detector < max pop size

Stage 2 : Tolerization period (T number generation)
Stage 1 process continues with new set of self (each loop)
until generations reaches T
If Generation = T
    Immature detector age = T (i.e born at generation
    1) becomes matured
Until # of immature + mature detector < max pop size

Stage 3 : At generation T + 1
For every antigen sets
    Present new antigen set to mature detector to be
    monitored
    If match, increment the match_count
Until all antigens are compared
If match_count > activation threshold
    If costimulation = True
        Mature Detector becomes Memory Detector
    else
        Delete mature detector
else (match_count < activation threshold)
    If age > life_span
        Delete mature detector
    Mature detector stays in the population

Stage 4 : At generation T + 2
Monitoring is done using memory detector
If match, delete antigen
The rest of antigens are presented to mature detectors
Repeat Stage 3
The antigens are then presented to immature detectors to
perform Negative Selection

Repeat Stage 4
```

Unlike the work of Hofmeyr that set the tolerization period in term of days, her tolerization period (*T*) is measured based on generation number and match count and the value of *T* is arbitrary. She also modified the DynamiCS to include gene library evolution and found out that with simulation of gene library evolution has reduced the amount of co-stimulation needed.

4.3 The work of Multi-shape Genes in IDS at Xi'an Jiaotong University, China [23]

The work to be presented here is concerning the genes evolution. Lu *et al.* [23] proposed a novel method for generation and evolution of gene libraries that can lead to effective Network-based IDS. His framework is illustrated below :

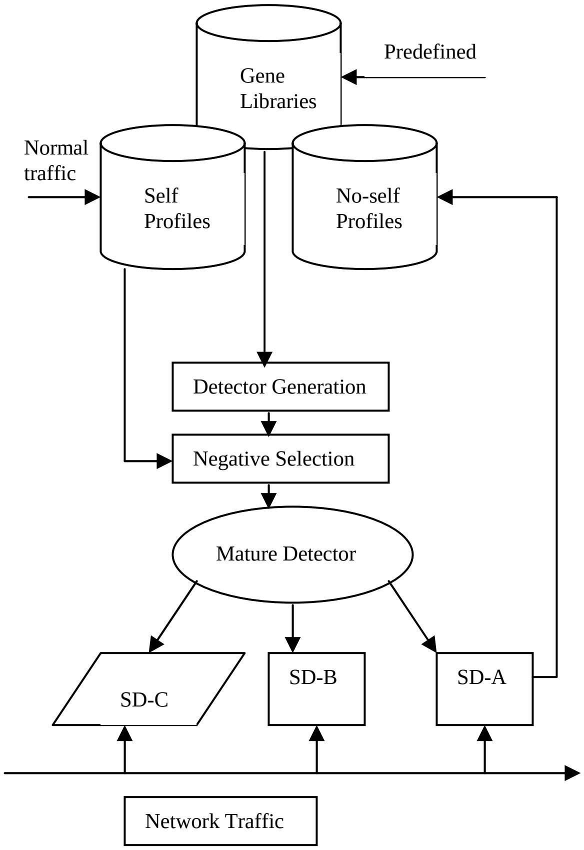

Figure 2 : Logic Architecture of the Artificial Immune Model [23].

The adaptive element in his work was the Detector Generator (DG) part where it generated profile and detector sets and its operational flow is listed below :

i) Gene Libraries Creation
   Build gene shape to describe network traffic pattern. Monitor normal traffic and construct profiles.
ii) Detector Generation
   Generate immature detector using randomly selected gene fragments from gene library
iii) Negative Selection
   Immature detectors that matched self profiles were deleted. Others became mature.
   Item (ii) and (iii) were repeated until the population reached a certain value.
iv) Gene Library Evolution
   When mature detector identified anomalous pattern, no-self profiles would be updated and transferred to hosts as memory detector (for misuse detection). Otherwise, the mature

detectors after exceeded age limit would be removed and new detectors would be generated.

The construction of genes was based on mechanism of the network protocol and their security holes. Since network attacks display different characteristics, genes were categorized into different categories. They are; Genes for individual TCP connection, Genes for grouped TCP connection and Genes for Telnet service. Both Forrest *et al.* [14] and Kim [1] also used features of the TCP connection as genes. Though the work presented here was another variation of approaches in addressing adaptability issue, the author make an attempt to cater different types of attacks by studying the different characteristics of attacks and had different types of genes. The authors indicated that the detection rate of the two former genes were promising. Unfortunately, no detailed result was given.

5. Summary

The discussion presented above has provided a brief overview of an adaptive IDS and three AIS algorithms, that represent three major evolutionary processes in human immune system which are negative selection, clonal selection and gene evolution. These processes contribute to learning and adaptation. All the works explained above use the notion of distinguishing self and non-self, negative selection and immune memory. Their approaches, have made the IDS adaptive by successfully responding to continuously changing patterns in normal network traffic.

The replacement of the immature detectors with a new set detectors was done in periodic manner. Hofmeyr introduced two and seven days for detectors' apoptosis, meanwhile Kim introduced age count up till certain generation to become mature. Similar to Hofmeyr and Kim, Lu *et al*. [23] had also deployed a life span of the detectors in his work. All the three works used the activation of mature detectors as a determinant to convert them into memory detectors or signature based detector. In conclusion, a periodic replacement of the useless detectors is the core element that contribute to the notion of adaptability in their works.

6. References

[1] Kim, Jung Won. “Integrating Artificial Immune Algorithms for Intrusion Detection”. PhD Thesis. Dept of Computer Science, University College of London 2002.

[2] U. Aickelin, , J. Greensmith, and J. Twycross, “Immune System Approaches to Intrusion Detection – A Review”, ICARIS 2004. pp. 316-329.

[3] Ye, N. Emran, M., Chen, Q. and Vilbert, S. “Multivariate Statistical Analysis of Audit Trails for Host-based Intrusion Detection”. IEEE Transaction on Computers, Vol. 51, No. 7, 2002.

[4] Z. Zhang and H. Shen. “Application of Online Training SVMs for Real-Time Intrusion Detection with Different Considerations”. Journal of Computer Communications xx 2004, pp. 1-15.

[5] C. Zhang, J. Jiang and M. Kamel. “Intrusion Detection Using Hierarchical Neural Networks”. Pattern Recognition Letters 26. 2005, pp. 779-791.

[6] M. O. Depren, M. Topallar, E. Anarim and K. Ciliz. “An Intelligent Intrusion Detection System (IDS) for Anomaly and Misuse Detection in Computer Networks”. Expert Systems with Application 2005, pp. 1-10.

[7] Anderson, J.P., ”Computer Security Threat Monitoring and Surveillance”, Technical Report, James P. Anderson Co., Fort Washington, PA, April, 1980.

[8] Denning, Dorothy E., “An Intrusion-Detection Model”, Proceedings of the Symposium on Security and Privacy. 1996. pp 118-131.

[9] Mykerjee, B., Heberlein, L. T. and Levitt, K. N., “Network Intrusion Detection”, IEEE Network, Vol. 8, No. 3, 1994, pp. 26-41.

[10] W. Lee, S.J. Stolfo and K.W. Mok, “Adaptive Intrusion Detection : A Data Mining Approach”. Artificial Intelligence Review 14: Issues on the Application of Data Mining. 2001 Kluwer Academic Publishers, Netherland. 2000, pp. 533-567.

[11] J. M. Tapiador, P. G. Teodoro and J. E.Verdejo, “Anomaly Detection Methods in Wired Networks : A Survey Taxanomy”, Journal of Computer Communications (2004) pp. 1-16.

[12] L.N. Castro, and J. Timmis, “Artificial Immune Systems : A New Computational Intelligence Approach”, Springer Verlag, London, Great Britain, 2002.

[13] L.N. Castro, and F. J. Zuben, “Artificial Immune Systems: Basic Theory and Applications. Technical Report TR-DCA 01/99, December 1999.

[14] S. Forrest, A.S. Perelson, L. Allen, R. and Cherukuri "Self-Nonself Discrimination in a Computer." S. Forrest,. In *Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy*, Los Alamitos, CA: IEEE Computer Society Press (1994).

[15] S. A. Hofmeyr, “An Immunological Model of Distributed Detection and Its Application to Computer Security”, PhD Thesis, Computer Scienec Dept of University of New Mexico, United States 1999.

[16] D. Dasgupta and F. Gonzales. “ An Immunity-based technique to characterize Intrusions in Computer Networks”. IEEE Transactions on Evolutionary Computation, 6(3) 2002. pp. 281-291.

[17] L.N. Castro and F. J. Zuben, “The clonal selection algorithm with engineering applications”. In Workshop Proceedings of GECCO’00, Workshop on Artificial Immune Systems and their Applications, LasVegas, USA, July 2000, pages 36–37.

[18] L.N. Castro and F.J. Zuben, “Learning and optimization using clonal selection principle. IEEE Transactions on Evolutionary Computation, Special Issue on Artificial Immune Systems, 6(3): 2001, pp.239–251.

[19] L. Sompayrac, How the Immune System Works, Blackwell Science Inc., 1999.

[20] M. Oprea and S. Forrest, “ How the Immune System Generates Diversity : Pathogen Space Coverage with Random and Evolved Antibody Libraries”, Proceeding of Genetic and Evolutionary Computation Conference (GECCO), July 1999. Available at http://www.cs.unm.edu/~forrest/ism_papers.htm

[21] M. Oprea, “Antibody Repertoires and Pathogen Recognition : The Role of Germline Diversity and Somatic Hypermutation, PhD Thesis, Department of Computer Science, The University of New Mexico, 1999.

[22] P. A. Vargas, L. N. Castro and F. J. Zuben, “Mapping Artificial Immune Systems into Learning Classifier”, 5th International Workshop, IWLCS 2002, Granada, Spain. Sept 7-8, 2002. pp. 163-186.

[23] J. Lu, B. Feng, B. Li and Y. Rao, “Study of a Multi-Shape-Gene Artificial Immune Model for Network Intrusion Detection”. Proceedings of the Second International Conference on MachineLearning and Cybernetics, Xi’an, 2-5 November, 2003.