IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE)
e-ISSN: 2278-1676,p-ISSN: 2320-3331, Volume 9, Issue 6 Ver. II (Nov – Dec. 2014), PP 20-30
www.iosrjournals.org

Artificial Immune System Research in the past years-an overview of existing methods, new initiatives and applications in intrusion detection systems

Sragdhara Bhattacharya
(M.tech 1st year,Department of Electrical Engineering,National Institute of Technology,Patna,India)

***Abstract*:** *The study of Artificial Immune Systems deals with the development of computational abstractions of the human immune system.AIS,in recent years has emerged as a new branch of Artificial Intelligence gaining ready acceptance in areas of computer security,anomaly detection,pattern recognition,robotics etc.There has been a lot of isolated work in this field till date which makes it difficult for anyone to have a complete overview of the evolution of AIS throughout the past years and its recent implementations.This review paper makes an attempt to present in a compact format,nearly all major works that have been done in the field of AIS and make a survey of existing methods and new initiatives.This paper focuses on the classical mechanisms of clonal selection,negative selection and immune networks theory,their limitations,the recent developments incorporating the dendritic cell algorithm,the new idea of "Danger Theory" that challenges the classical self-non-self viewpoint and their implementations in intrusion detection systems.*

Keywords: *Artificial Immune Systems,human immune system,clonal selection,negative selection,idiotypic networks,danger theory,dendritic cell algorithm,intrusion detection systems.*

I. Introduction

Different concepts in biology have always inspired the development of various effective computational models.The human immune system is a natural and competent system that very efficiently protects our body from a vast variety of foreign pathogens and hence is an inspiration for developing computational models and problem solving methods.The natural immune system has many interesting features-self-organised,self-monitored,error-tolerant,distributed,adaptive etc and also has powerful information processing abilities like feature extraction,pattern recognition,learning memory and so on.These are some qualities that are desired in all network security systems.Several applications of Artificial immune Systems have been made till date in fields concerning computer security or detection of faults and frauds.This paper basically makes a survey of the AIS theories in use since the time of its inception and its applications.

The Artificial Immune System(AIS),however exhibits a very high level of complexity.Different algorithms used in AIS actually imitate the behavior of the different types of cells of the immune system.In the first generation AIS,emphasis has been on Clonal Selection[1,2,3],Negative Selection[1,2,3] and Idiotypic approaches[1,2,3]which have often shown considerable limitations.So in recent years lot of developments have occurred and the second generation of AIS uses the Dendritic Cell Algorithm[4] and the Danger Theory[1,5]approach.AIS can certainly contribute with the already established methodologies in order to mutually improve their performances and application domains.

Section 2 gives a brief idea of the human immune system and the function of the cells of the innate and adaptive immune systems so that it becomes easier to understand the corresponding computational abstractions.Section 3 presents an overview of the first generation AIS algorithms namely clonal selection,negative selection and idiotypic approaches,their applications in anomaly detection and hence their limitations.Section 4 discusses the second generation AIS theories-the Danger theory and Dendritic cell algorithm(DCA) and their use in anomaly detection.This section also discusses the "niching" strategy of intrusion detection(clonal selection algorithm with embedded negative selection operator).Section 5 provides a brief discussion on the present day applications of AIS.Section 6 provides a conclusion about the essentiality of AIS and its future trends.

II. A Brief Overview Of The Natural Immune System

2.1 Innate immune system -The innate response is usually triggered when microbes are identified by pattern recognition receptors.Innate immune defenses are non-specific.Dendritic cells (DCs) are phagocytes in tissues that are in contact with the external environment and are located mainly in the skin, nose, lungs, stomach, and intestines.Dendritic cells serve as a link between the bodily tissues and the innate and adaptive immune systems, as they present antigen to T cells of the adaptive immune system.

2.2 Adaptive Immune System –It contains two types of cells(lymphocytes)-B-cells and T-cells both of which are derived from hematopoietic tissue of bone marrow.Both types of cells can recognize some specific targets.T-cells can recognize a particular "non-self" entity(pathogen) only when the antigens(fragments of the pathogen) are presented in combination with a "self" receptor called the major histocompatibility complex(MHC:antigen complex).Two types of T-cells are-killer T-cells and helper T-cells.Killer T-cells can recognize only the antigens in combination with MHC-I complex while Helper T-cells recognize only those antigens in combination with MHC-II complex.

- Killer T cells-They kill the virus-infected or damaged cells.Killer T-cells get activated when their T-cell receptor(TCR) binds with an antigen in a complex with MHC-I receptor of another cell.Another co-receptor CD8 on the T-cell aids the recognition of the MHC-I:antigen complex.This T-cell then travels throughout the body in search of other cells where the MHC-I receptors bear the same antigen.On coming in contact with such cells the T-cells release a cytotoxin name perforin that form pores in the target cell's membrane allowing the entry of water,ions and toxins.The entry of another toxin called granulysin(protease) induces the cell to undergo apoptosis(programmed cell death).
- Helper T cells- The helper T-cells have receptors that recognize only those antigens that are bound to MHC-II molecules.Recognition of MHC-II:antigen complex is aided by the helper T-cell co-receptors CD4 that recruits molecules within the helper T-cell that help in their activation.Activation of resting helper T-cells releases cytokines that enhance the microbicidal functions of macrophages and that of killer T-cells.Activation of helper T-cells causes an upregulation of molecules on the T-cell surface such as CD40 ligand that further activates antibody producing B-cells.

III. First Generation Ais Algorithms

3.1 Clonal selection

The clonal selection mechanism[1,2,3]proposed by Burnet in 1959 and it describes the B-cell cloning process as shown in Fig.1,basically says that only those cells that can recognize the antigens can proliferate and hence get selected over the others.The important features of this theory are-

- Firstly the B-cells are selected to be fit for the purpose during a "training period".Those B-cells that exhibit cell-surface receptors(antibodies) that match the corresponding antigen are selected to form the initial B-cell population while the rest that cannot bind to the antigen are removed.The selected B-cells are then released to the periphery.When they encounter the antigen,multiple versions of the B-cell receptors(antibodies) are produced that can bind to the matching antigen.
- The process of antibody tuning occurs through somatic hypermutation and affinity maturation[1].Somatic hypermutation ensures that the B-cell that exactly matches the antigen does not produce exact clones of itself that is the clones are slight variants of the parent cells.This is a kind of biological optimization that ultimately results in the production of antibodies that can bind more successfully to the antigens and hence may be used as detectors of non-self entities in the body.This whole process is called affinity maturation that produces the most responsive antibodies[1].
- When clonal selection operates on B-cells these can differentiate into long-lived memory cells[2] so as to encounter the repeated attacks by the same antigen during the entire lifetime of the individual.The initial exposure to an antigen induces an adaptive immune response that is initially handled by a small number of B-cells that produce antibodies of varying affinities.If we can store some high affinity antibody producing cells from the first infection so as to form a large initial specific B-cell sub-population(clone),then the effectiveness and speed of protection would be much more during later attacks[2].

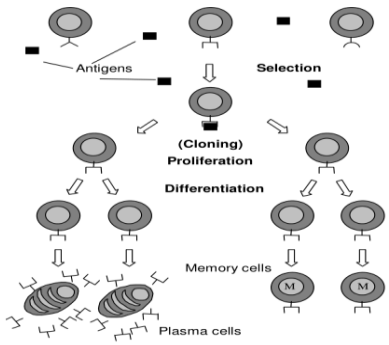


Figure 1. The clonal selection mechanism

Computationally the mechanism of clonal selection forms the basis of several population based algorithms.Many approaches in the literature have focused on optimization approaches like CLONALG[2].Also

this forms the basis for supervised learning algorithms like AIRS[2]. Application of the clonal selection theory leads to algorithms that evolve (through a cloning, mutation and selection phase), candidate solutions in terms of optimisation,or pattern detectors in terms of learning.In these algorithms we have populations of B-cells (candidate solutions) that match against antigens (functions to be optimised).These B-cells then undergo cloning (usually in proportion to the strength of the match) and mutation (usually,inversely proportional to the strength of the match). High affinity B-cells are then selected to remain in the population,Some low affinity cells are removed and new random cells are generated. Through this process,good solutions can be found, and in terms of dynamic environment and these solutions can be maintained over long periods of time.The clonal selection algorithm is represented as in Fig.4.

Figure 2. Clonal selection algorithm

3.2 Negative Selection Approach

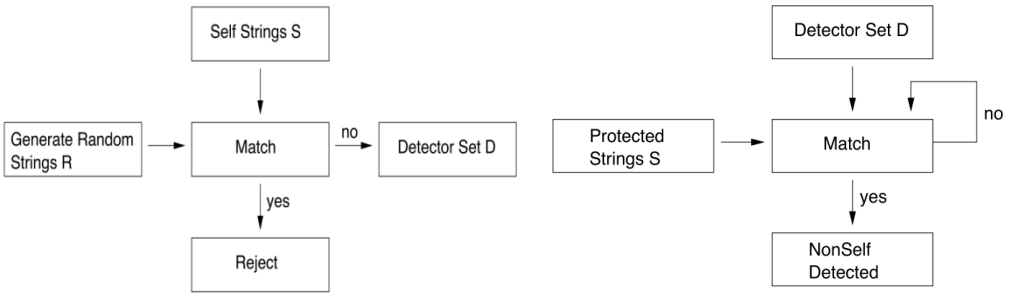
This theory was proposed by Joshua Lederberg in 1959.He suggested that when T-cells(another class of lymphocytes) are produced,they undergo a period of immaturity during which antigen recognition leads to their death that is the T-cells need further activation in the tissues to develop the ability to remove pathogens like bacterial agents and virus affected cells.The negative selection basically provides tolerance to the self cells that is it is the ability of our body to detect only the unknown antigens without reacting against the self cells.During the stage of embryonic development the T-cells migrate to a immune system organ called thymus till puberty.During this time the T-cells are exposed to a comprehensive version of self-antigens that is they undergo a pseudo-random gene rearrangement procedure followed by a censoring process in the thymus.During this time all T-cells having receptors matching to the self-antigens are removed by a filtering process.After puberty the thymus shrinks to a negligible size and hence all the self-reactive T-cells are eliminated.The matured T-cells are circulated in the lymphatic system now for protection purposes.

The negative selection algorithm was proposed by Forrest et al. in 1994[3].It was inspired by the ability of the immune system to train the T-cells to recognize only the antigens(non-self) and to prevent them from detecting the body's own cells(self).Basically we generate a set of (binary)detectors by randomly making candidates and then reject those that discard training self-data.A major application of a true NSA was in a system called 'Lisys'[1] which had three steps-

- Detecting self(a set **S** of elements of length **l** in shape space)
- Generating detectors(a set **D** of detectors such that no element of **D** matches any of the elements of **S**)
- Monitoring the occurrence of anomalies/anomaly detection(monitor a continual data stream by continuously matching **D** against the data stream)

The primary applications of NSA(as shown in Fig.3[2])is in the field of anomaly detection,self-non-self discrimination or pattern recognition where the detectors are generated in a complement space and they can detect changes in data patterns.A matching rule is chosen to establish similarity between two patterns and hence classify the self/non-self samples.The time complexity of generating detectors varies linearly with the size of self data.So we need to first estimate a size for the detector set for detecting anomaly with a certain reliability.If the threshold is too high,then the generated detectors become sensitive to any anomaly in data patterns and so

large number of detectors will be needed to provide a certain level of overall reliability.On the other hand, for a very low threshold value the available self set is not enough to generate the detectors.



Monitoring phase **Detecting Phase**

Figure 3. Negative selection algorithm as proposed by Forrest

3.2.1 Anomaly detection using Negative Selection

This approach was first proposed by Forrest et al. and extended by Forrest and Hofmeyr.This work is concerned with the detection of non-self in the area of network security where "non-self" refers to an undesired connection.All connections are modeled as "binary strings" and there is a pre-defined or known set of good and bad connections that is used to train and evaluate the algorithm.To construct an Artificial Immune System we need to create random binary strings called detectors.These detectors are made to pass through a maturation phase where these are presented with good i.e "self" connections.Those detectors that match any of these "self" connections will be eliminated.The remaining detectors mature but are not yet activated.If further during their lifetime they match anything else above a certain threshold value then they become activated.It is reported to a human operator who decides whether there is truly any anomaly.If there is a true anomaly then these detectors are promoted to "memory detectors" with an indefinite life span and minimum activation threshold[1,2,3,8].

An alternative approach to negative selection is the "positive-selection" approach used by Forrest et al. and also by Somayaji and Forrest, where "detectors for self" are evolved. A suspect non-self string would have to be compared with all self-detectors to establish that it is non-self, whilst with negative selection the first matching detector would stop the comparison.

3.2.2 Limitations of Anomaly Detection with Negative selection

- In negative selection inadequate detectors will result in false negatives(missed intrusions).
- Scalability is a disadvantage of the negative selection algorithm, as proposed by Kim and Bentley.As the systems to be protected become larger the sets of "self" and "nonself" also get larger.So it becomes very difficult to formulate a set of detectors that may provide adequate coverage and computational efficiency at the same time.
- The system needs to be infected above a certain threshold value for evoking a response from the AIS and also needs the interference of a human operator.This further causes a time delay increasing the time of exposure of the system to the damage.

3.3 Idiotypic networks(Immune Network Theory)

Jerne in 1974,postulated that the interactions between the antibodies(and not necessarily the external agents)cause modulations in the immune system as a whole,that leads to the generation of immune memory that is the ability to remember past encounters with pathogens so that the secondary response is quick and accurate[1].This theory proposes that antibodies can influence other antibodies as well as antigens.The clonal selection theory states that B-cells whose receptors show highest match to a stimulating antigen's binding region or epitope mature into plasma cells and are released into the blood stream following which the antibody binding sites called paratopes bind to antigen epitopes thus eliminating them.But according to Jerne's theory,an antibody also possesses a set of epitopes due to which they can be recognized by other antibodies as well.Epitopes that are unique to a certain type of antibodies are called idiotopes and group of antibodies sharing the same idiotope are said to belong to the same idiotype.If the idioptopes of an antibody are recognized by the paratopes of other antibodies then it is suppressed and its concentration is reduced while if the paratope of an antibody recognizes the epitope of an antigen then it is stimulated and its concentration increased.According to this theory the B-cells are not isolated but are continually communicating via collective dynamic network interactions,thus continually adapting the network.They communicate even in the absence of antigens.A system named "Jisys" is an example of an AIS based system inspired by Jerne's theory,developed by Hunt and Cooke and later by Timmis.It was developed based on the idiotypic network model formalized by Farmer et al. and later by Bersini and Varela et al. and it used the concept of stimulation and suppression effects within a network of antibodies.

The immune network theory may be utilized to control large populations of robots so that they have self-organising group behavior[3].B-cells are used to represent the robots.Each robot carries with itself the record of its success in collecting food while the other robots compare their success and strategies and stimulate or suppress the others accordingly.

IV. Second Generation Ais Approaches

4.1 Three main theories have challenged and augmented the process of self-non-self discrimination[1]-

- Costimulation
- Infectious non-self
- Danger Theory

Some of the cells that take part in the above theories belong to the innate immune system.

The Theory of Co-stimulation[1]came up when a problem was observed during the process of hypermutation of antibodies during clonal selection.If the hypermutated antibodies have a structure that can react to self antigens(self cells) then it would be an undesirable situation.So it was suggested that "B-cells" function in cooperation with "helper T-cells" that is B-cells would be eliminated only if they receive a costimulatory signal from the helper T-cells.In fact the helper T-cells are also regulated by a "stimulation cell" that provides the "costimulatory signal".These "professional antigen presenting cells" are called "dendritic cells(DC)"[1].

The Infectious non-self theory[1]was proposed by Janeway in 1989.She suggested that the dendritic cells(DCs)perform their own version of self-non-self discrimination.They can recognize the signatures of bacterial presence innately which is a skill that they have developed over millennia throughout the evolution of the species.The dendritic cells possess a repertoire of receptors on their surface for binding to molecules produced by the bacteria.These molecules are called PAMPs(pathogen-associated molecular patterns). Janeway showed that the induction of an immune response is facilitated by the production of costimulatory molecules from DCs. When exposed to PAMPs and antigen, the DC produces a collection of molecules that assist in their binding to a T-cell, increasing the time of contact of a T-cell with a presented antigen that helps in the activation of a T-cell.Infectious nonself can explain the need to add adjuvants to vaccines.

Even after the theories of costimulation and infectious self a confusion remained as to why the bacteria in our gut is not eradicated despite producing PAMPs and why at times the immune system responds to self.So finally the Danger Theory was proposed by Matzinger that proposed that the immune system responds by the detection of damage to the body and not by the detection of specific antigen structures or bacterial agents.

4.2 The Danger Theory and its applications in Artificial Immune Systems

This theory was proposed by immunologist Polly Matzinger in 1994.Some issues that provoked the development of the Danger theory[5] are:

- There is no immune reaction to the foreign bacteria in our gut or to the food we eat although these are non-self entities.
- Some autoreactive processes(reaction against self) are useful such as reaction against stressed cells.Also the immune system fights back tumours(reaction against self) while it supports successful transplants(no reaction to non-self).
- The definition of "self" is what is seen by lymphocytes during maturation.But the human body changes with time.So does the definition of "self".Thus defences against "non-self" learned early in life may become autoreactive at a later stage.

Matzinger concluded that the danger theory actually discriminates "some self from some non-self".According to this theory,it is not necessary for the immune system to attack everything that is foreign that is the immune system responds not simply to something that is non-self but to danger.In the Danger theory we take care of "self but harmful" and "non-self but harmless" invaders of our system.Here,danger is measured as damage to the cells indicated by distress signals sent out by cells when they die un-natural death(necrosis) unlike programmed cell death(apoptosis).A cell in distress sends out an alarm signal.The antigens in the neighbourhood are picked by the antigen presenting cells called macrophages which then migrate to the lymph node thus presenting the antigens to the lymphocytes.So the danger signal creates a "danger zone" around itself and all B-cells,producing antibodies,that match antigens in the 'danger zone' get stimulated and undergo the clonal expansion process.Those that are too far or do not match do not get stimulated.

The dendritic cells(DCs) of the innate immune system are responsive to both signals of necrosis and apoptosis as well as PAMPs[4].These are attracted to the areas where cells are dying from there they collect debris and all potential antigens.If a DC is exposed to the molecules from necrosing cells then the DC becomes mature.If it is exposed to the suppressing signals from apoptosing cells then the DC becomes semi-mature.Now the DC forms a complex with a T-cell that is the DC binds with a T-cell if the antigen collected and presented to the T-cell by the DC has enough affinity to the T-cell antigen receptor.If the DC is in a mature state,the T-cell

becomes activated thus eliminating the antigen from all entities and if the DC is semi-mature then the T-cell is not activated but is tolerised to the antigen and there is no response.

The following diagrams in Fig.4[5],have been proposed by Matzinger which show how Danger theory[5] may be viewed as an extension of immune signals or as an extension of "two-signal model" by Bretscher and Cohn where signal 1 represents antigen recognition and signal 2 represents costimulation.In Danger theory costimulation signal means "the antigen is really dangerous".In the first diagram,Burnet considers only signal 1 between infectious agents and lymphocytes(B-cells and killer T-cells Tk).The signal 2 was introduced by Brtescher and Cohn(as in second diagram).Signal 2 is produced by helper T-cells(Th) when they receive signal 1 from B-cells.B-cells present antigens to the helper T-cells and wait for the T-cells to recognize the antigen following which the immune response is elicited.According to Lafferty and Cunningham the helper T-cells also need to be activated by signals 1 and 2,both from antigen presenting cells(APCs)(as in third diagram).Helper T-cells can receive signal 1 from both B-cells(only non-self or selected antigens presented) and other APCs(random antigens presented) but signal 2 should be provided only for non-self antigens.Janeway proposed the idea of "infectious non-self" which allows the "priming" of APCs causing the production of signal 2.The priming signal is labeled as signal 0(as in fourth diagram).Matzinger proposes that the APCs are primed by the danger signal(as shown in fifth diagram) and also suggests that the efficacy of helper T-cells may be improved by routing signal 2 through APCs-labelling this signal as signal 3(as in sixth diagram).According to Matzinger "it is not necessary for the helper T-cells to see the same antigen as the killer T-cells but that the antigen should be presented by the same APC".

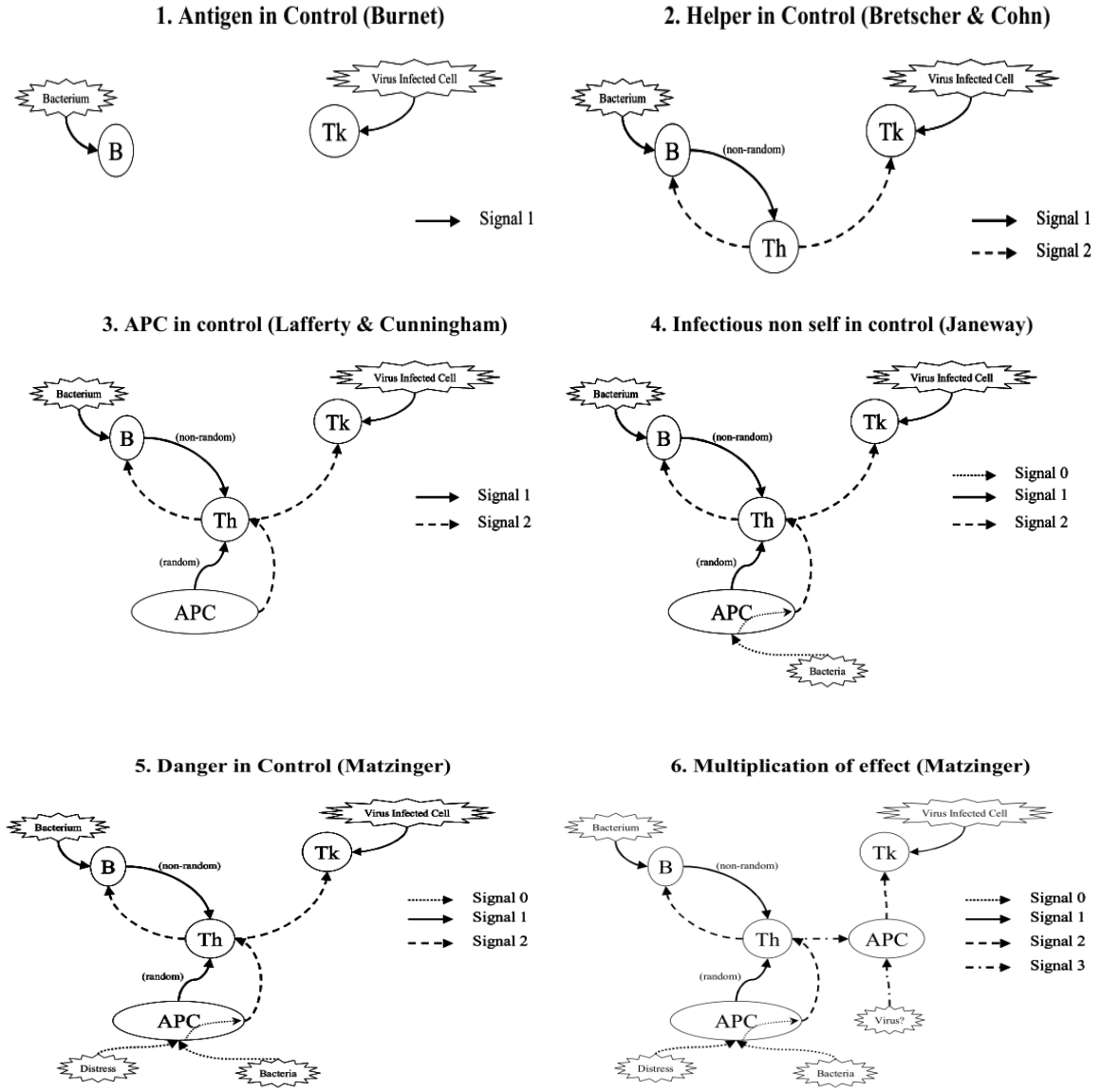


Figure 4. Danger theory viewed as an extension of immune signals

4.2.1 Anomaly Detection using Danger Theory

Self and non-self usually evolve and change during the lifetime of a system.So a system should be robust and flexible to continuously cope up with these changes[4].As the negative selection algorithm has certain limitations so we may use concepts from the danger theory to build an AIS that will not need to map self and non-self.Thus self-non-self discrimination is no more essential as now it is not the non-self entity but the danger signal that triggers an immune response.Once the danger signal is transmitted,the immune system can react to those antigens that is executables or connections that are "near" to the emitter of the danger signal.Here "near" does not refer to the physical nearness,but something that is meaningful in the context of network connections or IP addresses.

4.3 The Dendriric Cell algorithm

The DC biology is the backbone for the DCA[4].The DCs perform the sampling function in two areas.Input signal processing and antigen collection occur in the 'tissue'.After maturation the DCs migrate to the lymph node where they present antigens coupled with 'context' signals which are then interpreted and converted to immune signals.DCs exist in three different states-mature,semi-mature and immature.Transition of DCs to the mature state depends on the receipt of input signals.The resultant DC state is determined by the relative proportions of input signals received by the immature DC.The context of the antigen(that is the state of the signal environment) is decided by the terminal state of differentiation."Semi-mature" state implies a "safe" context while "mature" state implies a "dangerous" context.The collection of antigen forms by an immature DC is vital for the functioning of this system and forms the basis for the sampling function of the DCA.Each DC collects a subset of the total antigen set available for sampling.Dendritic cells respond to the changes in concentration of different molecules in their tissue environment.Safe signals initiate maturation upto the semi-mature state while danger and PAMP signals initiate maturation upto the fully mature state.Simultaneous receipt of signals from all different classes cause the increase in all three output signals,though the safe signal decreases the amount of IL-12 produced in response to the danger and PAMP signals.The amount of output signals is proportional to the concentrations of input signals received.

The DCA gives information about "how anomalous",a group of antigen is and not simply whether the antigen is anomalous or not.The DCA helps to "correlate" the otherwise incompatible data in the form of antigen and the signals and finally labels the antigens as "normal" or "anomalous".For this an "anomaly coefficient value"[4].is generated.The labeling of an antigen with a MCAV coefficient means correlating a series of input signals and the antigens.Here we use pre-normalised and pre-categorised input signals characterizing the data source.The four types of input signals are-PAMP,danger,safe signals and inflammation.

The abstract model of DCs must be transformed to an artificial immune algorithm to be of substantial use to AIS researchers.To transform it into an algorithm it should be formalized into a generic algorithm and a series of logical processes.All DC based algorithms are population based,the population consisting of a set of interacting objects,each object represented by a DC.Each cell in the population has a set of instructions that it performs each time it is updated.Each DC within the population performs its own version of antigen sampling and signal collection.Diversity and feedback is maintained within the population by the migration of DCs at different points of time that creates a variable "time-window" effect thus making the system more robust.Thus each DC in the population is an object with its own set of behavioral instructions.DCs process the input signals and then create cumulatively updated output signals,in addition to antigen collection during the entire sampling period.Each DC can exist in the immature,semi-mature and fully mature states.The initiation of state transformation from immature to either semi-mature or mature is not controlled by the collection of antigens but by exposure to the signals.Now the exposure of the DCs to the signals is again limited by the "migration threshold"[4].Immature DCs perform three main functions each time it is updated-

- Antigen sampling-The DC collects antigen from external sources(tissues) and stores them in the antigen storage data structure.
- Update input signals-The DCs collects values of all input signals present in the storage area.
- Calculation of cumulatively updated output signals-Each DC,receives input signals and produces the interim output signals which get added to form cumulatively updated output signals.

Signal processing occurs only in immature DCs.The signal processing equation is represented as a weighted sum equation[4]:

$$\text{Output} = \left(\text{PW} \sum_{i} \text{P}_i + \text{DW} \sum_{i} \text{D}_i + \text{SW} \sum_{i} \text{S}_i \right) * (1 + \text{I})$$

Where PW are the PAMP related weights,DW are for danger signals and SW are for safe signals whereas I stands for inflammation.P_i,D_i and S_i are the input signal values belonging to the categories of PAMP signals,danger signals and safe signals assuming that there are several input signals per category.'I' is the inflammation signal that appears three times-once for each output.The signal processing equation calculates the

interim output signal values for the CSM,semi-mature and mature outputs.These values are then summed up to produce cumulative values of signals.In the above equation the values of the PAMP weights are used to create all other weights relative to the PAMP weights.The two main functions of the three output signals are to (a)limit the life-span or sampling interval of the DCs and (b)to determine the terminal state of the DCs(semi-mature or mature) to find out whether the antigen is anomalous or not.

So basically in DCA multiple dendritic cells are used to create a population where each DC samples a set of signals within a given "time-window".Each DC is assigned a "migration-threshold" on its creation.As the cumulative output signals are updated,the DC compares the value it contains for CSMs with the value of its migration threshold.If the value of the CSM exceeds the migration threshold then that DC is removed sampling area and its life span terminated. Once the cell has migrated,it presents the antigens and output signals that it has collected throughout its life span.During this entire process,all kinds of signal that the DC has been exposed to throughout its life span are assessed and transformed into a binary value called the DC context.This is done by comparing the remaining two output signals-semi mature output(equivalent to interleukin-10) and mature output(equivalent to interleukin-12)using algorithm 2[4] as shown in fig 4.Pseudocode of the functioning of a generic DC object is shown in algorithm 1[4].

Algorithm 1:Pseudocode of the functioning of a generic DC object

```
input:signals of all categories and foreign entity
output:detection of foreign entity and context values
initialize DC;
while CSM output signal<migration threshold do
        get entity;
        store entity;
        get signals;
        calculate interim output signals;
        update cumulative output signals;
end
cell migrated to lymph node;
if semi-mature output>mature output then
        cell context 0;
else
        cell context 1;
end
kill cell;
replace cell in population;
```

Algorithm 2:Context assessment for a single DC

```
input:semi-mature and mature cumulative output signals
output:collected foreign entity and cell context
if semi-mature output>mature output then
        cell context 0;
else
        cell context 1;
end
print collected foreign entity plus cell context
```

Figure 5. Algorithm showing pseudocode for a generic DC object and Algorithm for context assessment for a single DC

4.4 Network Intrusion Detection by Clonal Selection Algorithm with a negative selection operator embedded in it-"Niching"

Just as the natural immune system protects the human body from a wide range of foreign pathogens,several computer based immune systems also can be used for the purpose of fault diagnosis,fraud detection and information security.[8].Intrusion detection is an area of vigorous research where the applicaton of Artificial Immune Systems has been examined(Dasgupta 1998,Somayaji 1997 et al).Intrusion Detection refers to the detection of malicious uses and unauthorized access of computer systems by both system intruders and external invaders.Such a software or device continuously inspects a system,points out malicious uses or violation of rules and reports to a management station.Several initiatives have been taken till date,to build Intrusion Detection Systems(IDS) using AIS(Mykerjee et al.).Kim and Bentley suggested that during the design of an artificial immune model for network intrusion detection the three design goals of "self-organised","distributed" and "lightweight"[8],should be satisfied and they identified such features of the natural human immune system,the adoption of which would help to meet these design specifications[8].Previously Forrest and Hofmeyr et al. had also applied the negative selection algorithm for building IDSs.But the use of the negative selection approach alone shows the problem of scalability that is, it works successfully only when network traffic data is less.The detection of various types of network intrusions requires the examination of huge amounts of network traffic data but the "random search feature" of the negative selection mechanism does not allow that and hence is not suitable for a large number of IDSs.So Kim and Bentley proposed the "niching" strategy whereby a clonal selection algorithm with a negative selection operator embedded inside it is used for network intrusion detection.It has three evolutionary stages- negative selection,clonal selection and gene library evolution[8].

Part of the AIS designed for network intrusion detection	**Biological equivalent**	**Function**
Primary IDS	Bone marrow/thymus	Generates numerous detector sets each of which describes abnormal patterns of network traffic packets and normal patterns of network traffic packets when network intrusions occur.This unique detector set is then transferred to the monitored single local host(secondary IDS).
Secondary IDS's	Monitored local hosts like lymph nodes	These receive the unique detector set from the primary IDS.
Network intrusions	Antigens	To attack the computer security system(any anomaly or intrusion).
Detectors	Antibodies	These serve as background processes for detecting whether non-self network traffic patterns are observed from the self network traffic patterns profiled at the monitored local host.

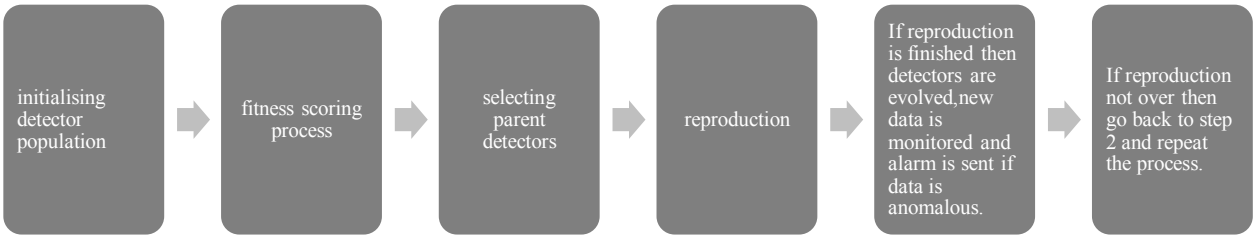


Figure 6. AIS algorithm for network intrusion detection

- Providing self and non-self antigens-As in the human immune system(where there are both "self" and "non-self" antigens)the input data to the AIS should also be classified into a self and a non-self set.Through the clonal selection procedure the antibodies will now evolve for the detection of non-self entities and leave the self entities as it is.
- Discretisation-Before passing on to the clonal selection module of the AIS,the self and non-self antigens should be processed through a discretiser.An antigen data has a large number of attributes,both continuous and discrete.But the continuous attributes represent a wide range of values while the detectors generated,use the binary genotypes.So discretisation is essential.
- Genotypes and Phenotypes-The detectors evolve gradually through clonal selection and they exist as a set of classification rules to classify self and non-self.In the 'if' part we put the combination of the different conditions to be tested and in the 'then' part we put the class label to be assigned to the rule.AIS employs binary genotypes.Each detector genotype has a conjunctive rule as its detector phenotype[8].The AIS basically creates an initial detector population by creating random genotypes each gene of which corresponds to an attribute of the detector phenotype..The total number of genes in the detector genotype is decided by the total number of attributes present in the antigen data.Each gene is made of nucleotides.The number of nucleotides again depends on the existing attribute values for example in Fig.7[8],gene 1/attribute 1 can have three possible values-TCP,UDP and TMCP.Each nucleotide is represented by a single binary bit,whose value of '1' denotes the inclusion of the corresponding attribute value in the 'if' part or 'condition' part of the classification rule while its value of '0' signifies its omission from the rule.Again if in a gene,the first bit is '1',then the following bits in that gene are read as '1' while if in the gene,the first bit is '0' then the remaining bits would be read as they are.
- Matching function-An attribute of a detector phenotype is represented by an interval having an upper and a lower bound while an attribute of an antigen phenotype has a single specific value.If we want to check whether a detector and an antigen matches them we need to check whether their attributes match that is if the specific antigen attribute value lies within any of the corresponding intervals of the detector attribute.
- Fitness scoring process-From the non-self antigen set,a sample set A of antigens is chosen.Again from the entire initial set P of detectors,a sample set D is chosen.The fitness value of each of the detectors in set D is initially set to 'zero'.Now each detector in set D is compared to the non-self antigens and the number of matching antigens for each detector is counted which is called the 'match-count' for that particular detector.The detector that shows the highest match-count has its fitness value increased by the 'match-count' while the fitness values of the other detectors remain the same.If more than one detector shows the

highest match-count then the fitness value is divided by the number of these at-par detectors and the fitness score is increased by that number.

- Reproduction and the negative selection operator-After the detectors in the detector population are evaluated,parent detectors are selected by the AIS to generate offspring detectors,the validity of which are again tested by the negative selection operator.M% worst detectors are substituted by N% best detectors(population overlapping).Parent detectors are selected by crossover and mutation of the two parents selected randomly from the N% fittest detectors.Any offspring that matches an antigen is discarded.
- Genetic operators used-In this clonal selection algorithm,two types of genetic operators have been used-crossover and mutation.Crossover occurs between genes or nucleotides that are the building blocks of the genotypes.Mutations render generality and specialisation to the detectors thus enhancing the process of pattern recognition.

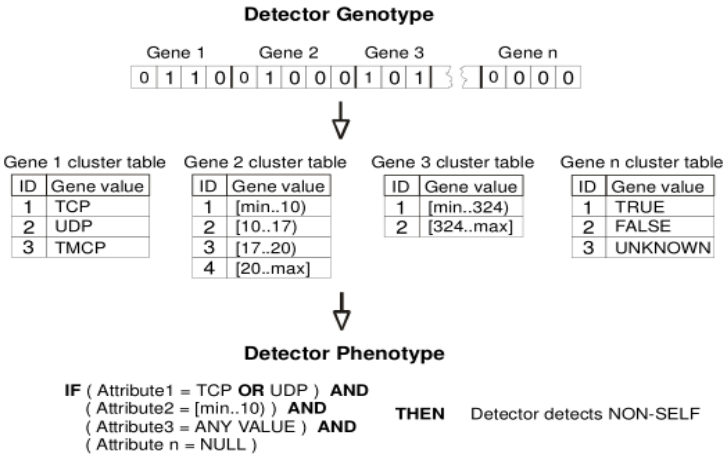


Figure 7. Genotype and Phenotype for intrusion detection model using AIS

V. Few Applications Of Ais

An artificial immune system can be used in error-detection system applied to robot based applications as well.Such applications have already been made for the immunisation of the fuzzy controller of a Khepera robot [9]that provides object avoidance and also for the control module of a BAE SYSTEMS RASCAL [9]robot.

In the Khepera robot system,the AIS learns normal behaviour(unsupervised)during a fault-free learning period and then identifies all errors greater than a preset error sensitivity.The fuzzy controller is used to run the Khepera robot around a maze.The robot's sensor data generates an appropriate motor speed that allows the robot to prevent avoid collisions with a walls.Each set of sensor values and motor speeds describe the robot's state.As each set of sensor values generates a known set of motor speeds therefore the data is deterministic and stable.Some sets of sensor values and motor speed values are not observed during the training period.These are called "non-self" data while all other instances of data are "self".The robot control system basically maps a single input to a single output in a deterministic manner with a given function.If this data describes a system state during normal behavior then any deviation from this data may be considered as anomalous.In the entire feature-space only the function represents "self-state" and the rest of the space is "non-self".A detector is said to match "non-self" if the system value is outside the range of the normal or self value.The second application of the AIS is to the control module of the BAE SYSTEMS RASCAL robot.The AIS is applied to one section of the control process used for this and other robots used by BAE systems-the Motion Control Manager(MCM)[9].

Artificial Immune Systems are also being used in detecting credit card frauds.AIS can detect all "non-standard" transactions without having been exposed to all such examples of fraudulent transactions during the training period.Here also the negative selection algorithm is implemented[10].The training data is preset at (0,1) and a predetermined number of detectors are created at random positions in the data space.During the training process any detector that falls within the radius of any member of the "self-set" is detected and rejected and replaced by another randomly generated detector.The process of detector generation iterates after generating a valid number of detectors.After this a new data vector is exposed to the population of detectors and if it does not lie within a hypersphere of radius of any of the self members then the new data vector is deemed to be non-self[10].Potentially,the detectors can take care of any instances of "non-self" even those not encountered before.This makes negative selection algorithm all the more apt for application in fraud detection as fraudsters are continuously devising novel methods of fraudulent transactions.

VI. Conclusion

This review paper has examined the topic of Artificial Immune Systems-the concept of abstraction of the natural human immune system and the different competing theories in the application and implementation of immune algorithms as exemplified by the DCA.The negative selection algorithm has even inspired the application of AIS in industries like in milling operations where this algorithm is used to detect tool breakage[11].Here self is defined as normal cutting operations while any deviation from the normal cutting force is deemed as "non-self".The DCA is performing quite well across a range of different problems as compared to the other nature-inspired techniques.AIS forms an excellent error detection system.Its application in detecting the errors injected in the Khepera robot fuzzy controller and MCM following the learning period is novel as the AIS used is independent of the system(depends only on the input and output) and does not require any information about the system except that the system should be stable and deterministic[9].AIS is distinct from other biologically inspired approaches in that here no archetypal approach is used.Several algorithms mimic the behavior of the different types of cells present in the natural immune system.The AIS algorithms are still evolving.Therefore the future is still uncertain but there is ample scope for development in this field.

References

[1]. Julie Greensmith,Amanda Whitbrook,Uwe Aickelin,Artificial Immune Systems
[2]. Jon Timmis,Artificial Immune Systems-Today and Tomorrow,April 7th,2006
[3]. D.Dasgupta,Z.Ji,F.Gonzalez,Artificial Immune System(AIS) research in the last 5 years
[4]. Julie Greensmith,The Dendritic Cell Algorithm,thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy,October 2007
[5]. Uwe Aickelin,Steve Cayzer,The Danger Theory and its applications to Artificial Immune Systems, Proceedings of the 1st International Conference on Artificial Immune Systems (ICARIS-2002), pp 141-148, Canterbury, UK, 2002.
[6]. L.N.de Castro,J Timmis,Artificial Immune Systems as a novel soft computing paradigm
[7]. Ayi Purbasari,Iping Supriana S,Oerip S Santoso,Rila Mandala,Designing Artificial Immune System based on Clonal Selection using agent-based modeling approach.
[8]. Jungwon Kim,Peter J. Bentley,The Artificial Immune System for Network Intrusion Detection:An Investigation of Clonal Selection with a Negative Selection Operator
[9]. Richard Canham,Alexander H Jackson,Andy Tyrrell,Robot error detection using an Artificial Immune System
[10]. Anthony Brabazon,Jane Cahill,Peter Keenan,Daniel Walsh,Identifying Online credit card fraud using Artificial Immune Systems
[11]. Dipankar Dasgupta,Stephanie Forrest,Artificial Immune Systems in industrial applications