World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:8, No:10, 2014

Apoptosis Inspired Intrusion Detection System

R. Sridevi, G. Jagajothi

***Abstract*—**Artificial Immune Systems (AIS), inspired by the human immune system, are algorithms and mechanisms which are self-adaptive and self-learning classifiers capable of recognizing and classifying by learning, long-term memory and association. Unlike other human system inspired techniques like genetic algorithms and neural networks, AIS includes a range of algorithms modeling on different immune mechanism of the body. In this paper, a mechanism of a human immune system based on apoptosis is adopted to build an Intrusion Detection System (IDS) to protect computer networks. Features are selected from network traffic using Fisher Score. Based on the selected features, the record/connection is classified as either an attack or normal traffic by the proposed methodology. Simulation results demonstrates that the proposed AIS based on apoptosis performs better than existing AIS for intrusion detection.

***Keywords*—**Apoptosis, Artificial Immune System (AIS), Fisher Score, KDD dataset, Network intrusion detection.

I. Introduction

SYSTEMS like firewalls or authentication mechanisms are no longer enough to provide security for existing network systems. Conventional intrusion detection systems integrate available information from a system that provides normal activity details called Self, to ensure that the system can differentiate unusual activities, categorized as Nonself. Unauthorized actions are detectable as an intruder's behavior will be different from that of a legitimate user. Hence a system attack can be determined by collecting and analyzing operating system and network activity data [1].

Intrusion Detection System (IDS) monitors continuously and record current activities which are then compared with past normal behavior. Based on the deviation level, it is then classified as attacks, errors or abnormalities. IDS do not react against attacks immediately. They generally inform the administrator about an intrusion through several attack detecting methods [2]. Monitoring/analyzing network activities, locating a network's vulnerable spots and integrity testing of sensitive and important data are some ways IDS operates to detect intrusions [3].

The IDS software analyzes/automates an intrusion detection process effectively [4], checking all possibilities surrounding network activity and identifies suspicious signatures that indicate a network/system attack. Being up-to-date on current affairs in computing can confirm various attacks on branded company servers. Though firewalls are a useful defense their current technology makes them insufficient in detecting/blocking all types of attacks [5].

Fig. 1 shows a Network Intrusion Detection in computer systems. An IDS can be located in a router, Firewall or LAN to detect network intrusions and provide information about known signatures. Thus, normal activity and intrusions classification accuracy has an important role in an IDS's efficiency.

Classification models are based on various algorithms and are the tools which partition given data sets into various classifications based on specific data features [6]. Currently, data mining algorithms are popular for effective classification of intrusion using algorithms including decision trees, naïve Bayesian classifier, neural network, and support vector machines. But classification accuracy of many present data mining algorithms should be improved as attack detection is difficult as attackers constantly change attack patterns. Network intrusion detection models that are presently used have high false positives. For efficiency, the intrusion detection model depends on Detection Rates (DR) and False Positives (FP). DR is the number of intrusion instances detected correctly by the system and divided by total intrusion instances in the analyzed data. FP is an alarm which indicates that an attack might not really be so. IDS aim to maximize DR and minimize FP.

Signature-based methods limitation is that they are unable to detect cyber threats; as such threats are launched through unknown attacks patterns. When a new attack is discovered and its signature is developed, it requires a large latency period to ensure its deployment on networks. Such shortcomings led to the need for discovery of intrusion detection techniques for identifying attacks. Artificial Immune Systems (AIS) are a relatively new research area having a lot of potential to solve various related issues [7]. Its growth has ensured many new techniques and approaches for solving problems.

AIS is increasingly popular today. Many immunology concepts have been used for solving real-world science and engineering problems like theoretical modelling and simulation in various applications [8]. The Human Immune System (HIS) is a rich theory source which inspires creation of new approaches to computational problems and so this new field is known as Immunological Computation (IC) [9].

R.Sridevi is with Shri Angalamman College of Engineering and Technology, Tiruchirappalli, Tamilnadu, India (Phone: +91 9066002047; e-mail: sridevi.odm@gmail.com).

G. Jagajothi is with Periyar Maniammai University, Tanjore, Tamilnadu, India (e-mail: gjagajothi@yahoo.in).

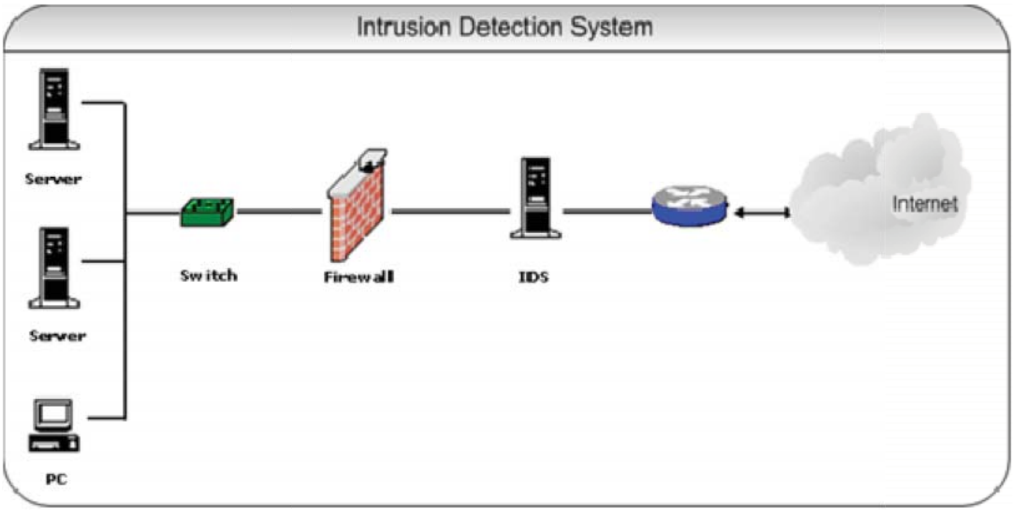

Fig. 1 Network with Intrusion Detection System

AIS are highly distributed, adaptive and self-organizing, with a memory of past events and having the ability to learn new encounters. It includes varied intelligent methodologies which in turn sprout effective solutions to real world problems [10]. HIS performs similar to an intrusion detection system in a computer system. Forrest et al. [11] first suggested that human immune concepts be applied to computer security. Studies reveal that the immune based IDS is formulated on human immune system concepts, performing similar tasks of innate and adaptive immunity. A normal behavior profile is generated by collating appropriate audit area services behavior. Rao et al. [12] dealing with various methodologies in intrusion detection suggested combining expert systems when performing various classification techniques like Bayesian decision, Neural Networks, and Fuzzy logic.

Zhao et al. [13] describe various research trends of immunity based network intrusion detection and refers to anomaly based intrusion detection attacks and its framework. Properties obtained by using this framework include diversity, distributivity, locality, adaptability and dynamicity.

This paper proposes an improved AIS classifier based on programmed cell death, also known as apoptosis, for intrusion detection. The proposed method is compared with AIS.

II. Materials and Methods

KDD'99 dataset was used to investigate the effectiveness of proposed AIS technique. Features are selected using Fisher Score and the obtained features classified using AIS and the proposed AIS classifier. Sections A, B and C explain the process in detail.

A. Dataset

Experiments were performed using a KDD '99 dataset with tests being conducted using ten-fold cross-validation. Four types of attack were detected using the proposed technique. The attack types being:

- Denial of Service (DoS),
- Remote to Local (R2L),
- User to Root (U2R) and
- Probe.

During DoS attacks, attackers send malformed or strange packets to the victim with the aim of disrupting service by trying to limit access to a machine/ service. Examples are back, land pod, teardrop and smurf. When the operating system services are compromised, attackers gain unauthorized access from a remote machine during R2L attack. Examples are Ftp_write, Guess passwd, Imap, warezclient, warezmaster, phf, spy and multihop.In U2R attacks, attacker gains unauthorized access to root privileges and on compromising, programs run code on behalf of the attacker (e.g., login as root, su to root, etc.). Examples are Load module, Perl, rookit and buffer overflow. In probing attacks, attacker scans the network to collect information or to locate vulnerabilities. Examples are ipsweep, nmap, port sweep and Satan [14].

B. Feature Selection – Fisher Score

Fisher score is one of the most widely used filter based, supervised feature selection method. The most relevant features for classification are determined using the Fisher score. The Fisher score is a supervised method with class labels and features with best discriminant ability is found [15]. If n_i is the number of samples in class *i*, μ_r^i and $\left(\sigma_r^i\right)^2$ is the mean and variance of class *i*, (i=1,…,c) for feature *r* the Fisher score is computed by (1):

$$F_r = \frac{\sum_{i=1}^{c} n_i \left(\mu_r^i - \mu_r\right)^2}{\sum_{i=1}^{c} n_i \left(\sigma_r^i\right)^2} \qquad (1)$$

High Fisher score indicates that the feature is more discriminative. A subset of features is selected using the Fisher score. However, it may lead to a suboptimal subset of features. To overcome this problem, a generalized Fisher score which jointly select features are used [16]. In generalized Fisher score, a subset of features is obtained by maximizing the lower bound of Fisher score. This results in a mixed integer programming, which is converted to a quadratically constrained linear programming. A cutting plane algorithm is utilized for finding solutions where in each iterationa multiple kernel learning problem is resolved alternatively by multivariate ridge regression and projected gradient descent [16]. To find a feature subset of size ‘m’ containing the most informative features using generalized Fisher score can be obtained by Fig. 2.

Notations:
C- number of classes;
Ω - a polynomial sized subset;
d - is the number of features;
p – indicator variable which represents where feature is selected or not;
P – p^t∈ P;V is a Lagrangian multiplier;
Λ-kernel weights;
γ I – perturbation term.

Pseudo code for Generalized Fisher Score for Feature Selection

```
Input: C and m;
Output: V and Ω;
Initialize $V = \frac{1}{n} 1_n 1_c^T$ and t = 1;
Find the most violated constraint $p^1$, and set $\Omega_1 = \{p^1\}$;
repeat
    Initialize $\lambda = \frac{i}{t} 1$
    repeat
        Solve for V under the current λ:
        $V = \left( \frac{1}{\gamma} \sum_{t=1}^{|P|} \lambda_t \sum_{j=1}^{d} p_j^t K_j + I \right)^{-1} H$
        Solve for λ using gradient descent as:
        $\nabla_{\lambda t} g(\lambda, V) = -\frac{1}{2\gamma} tr \left( V^T \sum_{j=1}^{d} p_j^t K_j V \right)$
    until converge
    Find the most violated constraint $p^{t+1}$ and set$\Omega_{t+1} = \Omega_t \cup p^{t+1}$;
    t = t + 1;
until converge
```

Fig. 2 Fischer Score based feature selection

C. Proposed Methodology

Immunologists believe that 'immunity is the identification of the Self and Nonself, and eliminating Nonself ensures the body’s integrity as a physiological response’. The skin, physiological conditions, congenital immune system and adaptive immune system form the human’s natural immune system. The HIS’s primary function can be said to be the differentiation between things which belong in the body and those that do not. The HIS can differentiate between self and non-self antigens and the process of detecting and removing non-self includes both innate and adaptive immunity. Innate components are nonspecific and unchanging in spite of repeated antigen exposure. Adaptive immunity is specific and involves memory which permits the immune system to react more quickly when an antigen is located the second time. Hall et al. evaluated immune based system by introducing architecture with two systems leading to promising results after system tests [17].

Associations between immune and computer systems are quite strong. The former safeguards the body from pathogens, similar to a computer security system protecting a computer from malicious users. AIS is now attracting a lot of attention in the monitoring of engineered systems. Processes of the natural immune system are applied to solve real world problems using AIS [18].

AIS technology attempts to model defense mechanism characteristics and functionalities of living beings. Such a defense mechanism allows an organism to safeguard against foreign substance invasions. Such substances recognition is based on a key and lock analogy, where the aim is the location of antibodies with the best immune response to an invading antigen [19].

The natural immune system’s genetic memory stores excellent antibodies which are later used to identify antigens which invaded the organism earlier. This in turn leads to a quicker response. The biological environment’s new functionalities were observed to model a new immunological approach, principally through the organization/clustering of similar antibodies (Ab) through the process. It is felt that such functionalities improve the AIS recognition capacity [20].A matching concept is used to search for a solution. As AIS are evolutionary algorithms, they suit problems that change over time requiring solutions repeatedly rather than being one-off optimizations [21].

In this paper, the Apoptosis mechanism of the immune system is used for classification. A cell can die in two ways; necrosis and apoptosis. The former is when a cell is damaged by an external force like poison, a body injury, infection or being cut off from a blood supply. Apoptosis is relatively civil; it’s a cell’s suicide. Apoptosis cleanup is easier, sometimes being referred to as programmed cell death as apoptosis follows a controlled, predictable routine.

When a cell kills itself, proteins called caspases start acting. They break down cellular components required for survival spurring production of enzymes called DNases, which in turn destroy the cell’s nucleus’ DNA. The cell shrinks sending out distress signals answered by macrophages which are known as vacuum cleaners. They clean the shrunken cells without a trace. Hence these cells are unable to cause damage similar to necrotic cells. Fig. 3 shows the apoptosis procedure.

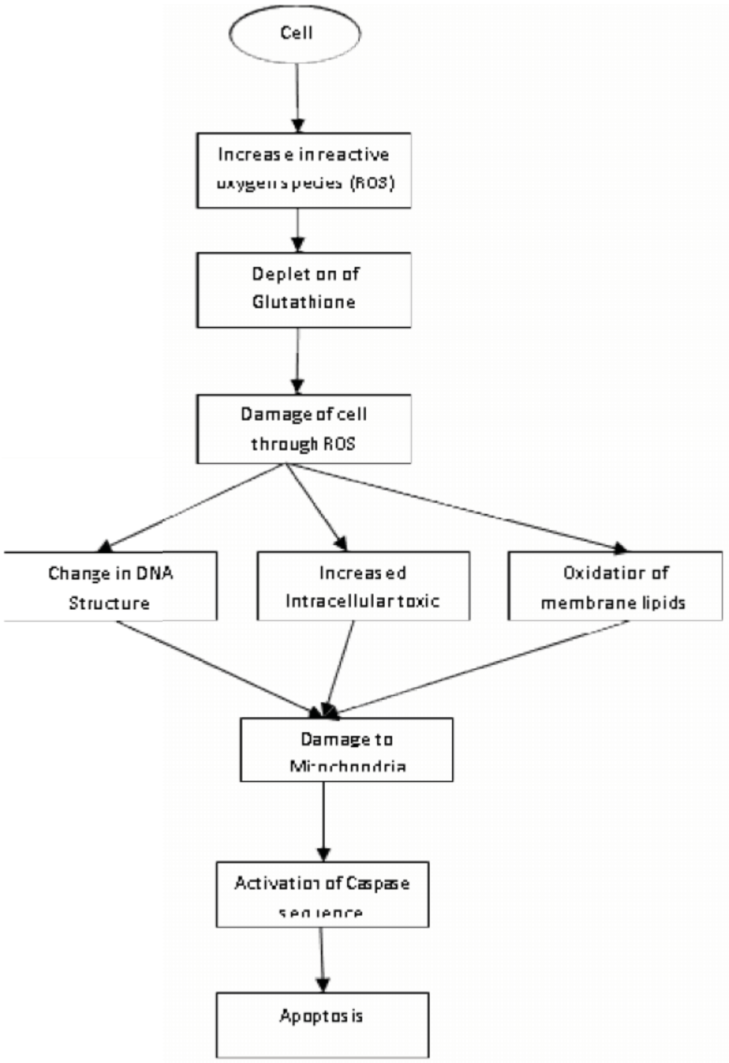
Cell
Increase in reactive oxygen species (ROS)
Depletion of Glutathione
Damage of cell through ROS
Change in DNA Structure
Increased Intracellular toxic
Oxidation of membrane lipids
Damage to Mitochondria
Activation of Caspase sequence
Apoptosis

Fig. 3 Apoptosis Process in Mammals

The proposed architecture of the Artificial Immune System classifier based on apoptosis is shown in Fig. 4. The features obtained using generalized Fisher Score are represented as input vector to the proposed AIS method for classification. In the initial stage of the proposed method, the input is pre-processed before AIS algorithms are applied. The input vector is checked for categorical features which are generalized. The input vector is normalized. Manhattan distance is used to measure affinity, and the affinity threshold is measured.

After the initial process, the AIS algorithms are applied. The artificial recognition balls (ARB's) are formed based on the affinity threshold computed and the best ARB is matched to the antigen. The elite antibodies are retained in ARB and the intra antibody affinity is computed. The antibodies with close affinity are mutated and matched with antigen. This match decides whether the parent or the child undergoes apoptosis. If the mutated antibody is fitter than the parent antibody then parent is subjected to apoptosis else the child is subjected to apoptosis. Thus, it ensures that the best antibody survives. The selected antibody is further mutated within the ARB. The ARB competition for resource is performed and now is checked for termination criteria. On reaching termination criteria, the algorithm ends else the ARB's are further subjected to generating and mutating. The elite from each ARB is selected and the antibodies are included in the memory cell. Before classifying with k-nearest neighbor, the antibodies are ranked based on usage and those with low score than the threshold are eliminated.

For an attack $n \in N$

The PCD is computed using the function

$$A = \eta + \sum_{i=1}^{N} p_{i*} G_i$$

where η is a constant with default value 0.5

p_i is the probability of a specific type of attack

G_i is the probability of the attack of a specific group

N is the number of known attacks

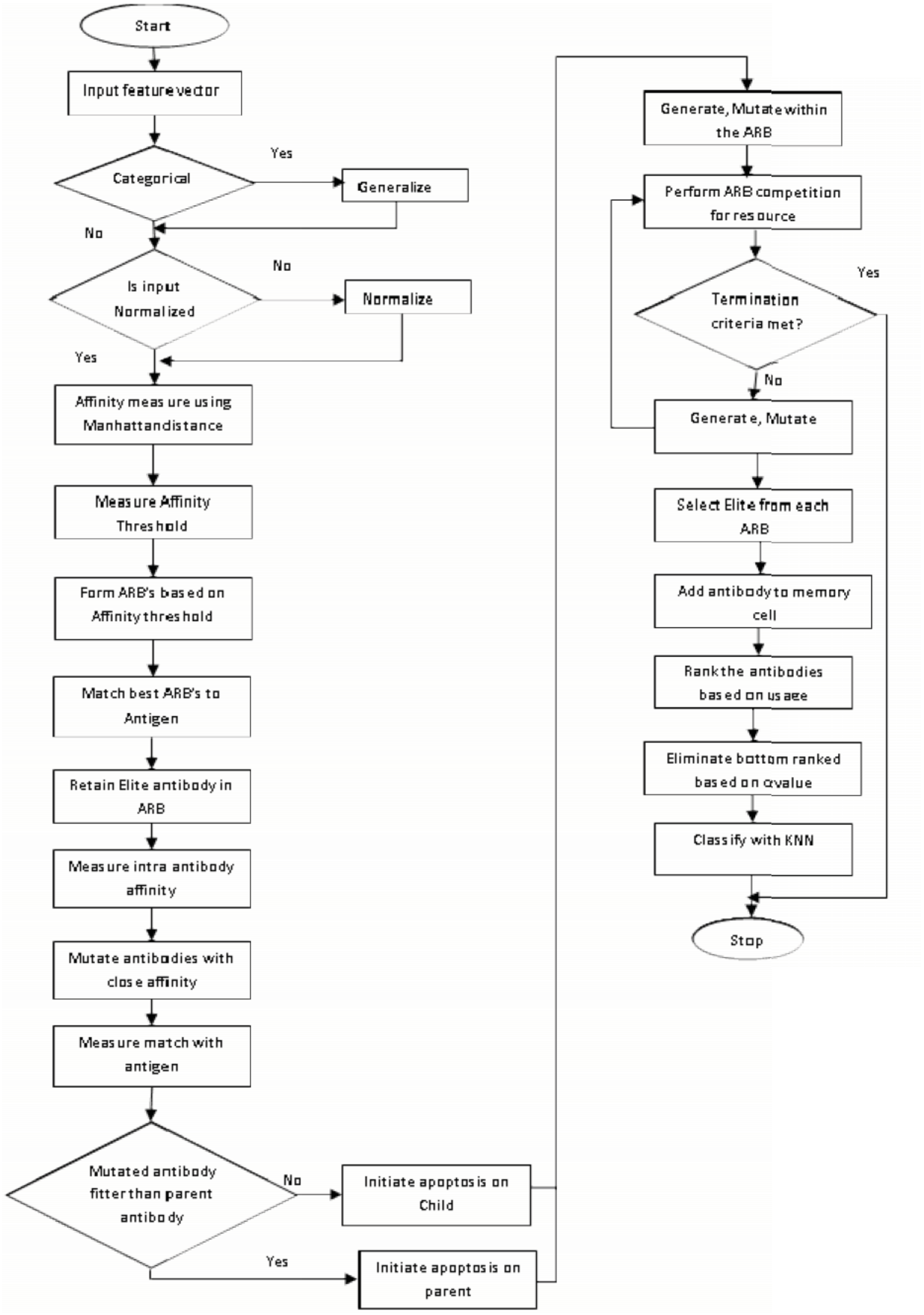
Start
Input feature vector
Categorical
Yes
Generalize
No
Is input Normalized
No
Normalize
Yes
Affinity measure using Manhattandistance
Measure Affinity Threshold
Form ARB's based on Affinity threshold
Match best ARB's to Antigen
Retain Elite antibody in ARB
Measure intra antibody affinity
Mutate antibodies with close affinity
Measure match with antigen
Mutated antibody fitter than parent antibody
No
Initiate apoptosis on Child
Yes
Initiate apoptosis on parent
Generate, Mutate within the ARB
Perform ARB competition for resource
Yes
Termination criteria met?
No
Generate, Mutate
Select Elite from each ARB
Add antibody to memory cell
Rank the antibodies based on usaee
Eliminate bottom ranked based on αvalue
Classify with KNN
Stop

Fig. 4 Flowchart of the Proposed Methodology

The fitness function is evaluated using

$$F(I) = \frac{2 * P * R}{(P + R)}$$

where P is the precision, R is the recall.

III. Results and Discussion

Experimental results are summarized to construct patterns for intrusion detection over the KDD'99 datasets. The evaluation of the proposed AIS classifier based on apoptosis was carried out by integrating our Java code withWEKA data-mining tool and its classification accuracy is compared with Artificial Immune Recognition system [22]. Table I shows the training summary of the AIS and the proposed method. Table II gives the results of the classification and Root Mean Squared Error (RMSE). Tables III and IV give the detailed accuracy by class and confusion matrix for AIS and the proposed method.

TABLE I
Training Summary of Various Methods

	AIS	AIS with PCA	AIS + Fisher score feature selection	Proposed AIS technique	Proposed AIS with PCA	Proposed AIS technique + Fisher score
Affinity Threshold	0.227	0.058	0.229	0.178	0.145	0.162
Mean ARB clones per refinement iteration	51.313	51.817	51.533	51.275	51.812	15.487
Mean memory cell clones per antigen	19.606	19.979	19.79	19.238	19.82	19.476

TABLE II
Summary of Classification Accuracy and RMSE

	AIS	AIS with PCA	AIS + Fisher score feature selection	Proposed AIS technique	Proposed AIS with PCA	Proposed AIS technique + Fisher score
Correctly Classified Instances	99.35%	99.54%	97.90%	99.37%	99.51%	99.65%
RMSE	0.138732	0.0479	0.1024	0.1073	0.0521	0.0428

The proposed technique improves the classification accuracy by 1.97% compared to a similar feature selection technique used by AIS. Table III shows the precision and Recall obtained using existing AIS under different feature selection techniques. However from Table IV it can be seen that precision and recall outrank existing AIS method establishing the proposed technique as a better solution for IDS.

TABLE III
Detailed Accuracy by Class for AIS

Class	Precision with AIS	Recall with AIS	Precision with AIS and PCA	Recall with AIS and PCA	Precision with AIS and Fisher Score	Recall with AIS and Fisher score
Normal.	0.992	0.994	0.997	0.998	0.999	0.979
Teardrop.	0.612	0.582	0.677	0.603	0.217	0.932
Satan.	0.987	0.987	0.987	0.987	1	0.987
Nmap.	0.963	1	0.963	1	0.938	0.962

TABLE IV
Detailed Accuracy by Class for Proposed AIS

Class	Precision-Proposed AIS technique	Recall-Proposed AIS technique	Precision -Proposed AIS with PCA	Recall-Proposed AIS technique	Precision-Proposed AIS technique + Fisher score	Recall-Proposed AIS technique + Fisher score
Normal.	0.997	0.992	0.997	1	1	1
Teardrop.	0.618	0.608	0.714	0.728	0.764	0.954
Satan.	0.976	0.989	0.992	0.992	0.992	0.994
Nmap.	0.921	1	1	1	1	1

IV. Conclusion

AIS protects a complex system against malicious defects achieving its efficiency through an extension of the concept of organizing multi-cellular organisms to information systems. AIS main features include self-maintenance, distributed and being adaptive to computational systems. The mechanism of the apoptosis based human immune system has been adapted to build an intrusion detection system for protecting computer networks. The proposed method is compared with AIS. Experiments reveal that accuracy, precision and recall of the proposed procedure is better than that of AIS.

References

[1] Hui Wang, Guoping Zhang, Huiguochen and Xueshu Jiang, "Mining Association Rules for Intrusion Detection",2009 IEEE International conference on frontier of Computer Science and Technology.

[2] ChristophEhret, Ulrich Ultes-Nitsche, Immune System Based Intrusion Detection System University of Fribourg Department of Computer Science, University of Fribourg,Boulevard de Pérolles 90, CH-1700 Fribourg, Switzerland.

[3] S. Northcutt and J. Novak, "Network Intrusion Detection:An Analyst's Handbook," 2nd Edition, New Riders Publishing,Berkeley, 2000.

[4] Karen Scarfone, Peter Mell, Guide to intrusion detection and prevention systems (IDPS) Special Publication 800-.94,2007

[5] L de Castro, J Timmis, Artificial Immune Systems: A New Computational Intelligence Approach, Springer Verlag, 2002.

[6] Sophia Kaplantzis, Nallasamy Mani, A Study on Classification Techniques for Network Intrusion Detection

[7] U. Aickelin and D. Dasgupta, Artificial Immune Systems Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques,2008.

[8] DipankarDasgupta, Artificial Immune Systems: A Bibliography CS Technical Report No. CS-07-004 December 2007 Version 5.8.

[9] John E. Hunt and Denise E. Cooke, Learning using an artificial immune system, Journal of Network and Computer Applications (1996) 19, 189–212 Ó 1996 Academic Press

[10] ChingthamTejbanta Singh, and Shivashankar B. Nair, An Artificial Immune System for a MultiAgent Robotics System, World Academy of Science, Engineering and Technology 11 2005

[11] S. Forrest, A. S. Perelson, L. Allen, and R. Cherukuri. Self-nonself discrimination in a computer. Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy, pages 202–212, Oakland, CA, 1994. IEEE Computer Society Press.

[12] Shaik Akbar, Dr. K. Nageswara Rao, Dr. J. A. Chandulal, Intrusion Detection System Methodologies Based on Data Analysis, International Journal of Computer Applications (0975 – 8887) Volume 5– No.2, August 2010

[13] Zhao junzhonghuanghoukuan , An evolving intrusion detection system based on natural immune system proceedings of IEEE TENCON'02

[14] Leandro N. de Castro and Jon Timmis(2002). An artificial immune network for multimodal function optimization. In IEEE Congress on Evolutionary Computation (CEC), pages 699–704.

[15] Gu, Q., & Han, J. (2011, October). Towards feature selection in network. In Proceedings of the 20th ACM international conference on Information and knowledge management (pp. 1175-1184). ACM.

[16] Gu, Q., Li, Z., & Han, J. (2012). Generalized fisher score for feature selection. arXiv preprint arXiv:1202.3725.

[17] John M. Hall,AN Investigation into Immune-Based Intrusion Detection, December 2003, University of Idaho.

[18] Kaushik Ghosh and Rajagopalan Srinivasan, Immune-System-Inspired Approach to Process Monitoring and Fault Diagnosis, Copyright © 2010 American Chemical Society.

[19] De Castro, L. N. &Timmis, J. I. (2002). Artificial Immune Systems: A Novel Paradigm for Pattern Recognition, In : Artificial Neural Networks in Pattern Recognition, L. Alonso, J. Corchado, C. Fyfe, 67-84, University of Paisley.

[20] K. Regina, A. Boukerche, J. Bosco, M. Notare, "Human Immune Anomaly and Misuse Based Detection for Computer System Operations: Part II", Proceedings of the International Parallel and Distributed Processing Symposium 2003, IEEE © 2003.

[21] Zhu, Dan , Data mining for network intrusion detection: A comparison of alternative methods Decision Sciences Date: Monday, October 1 2001.

[22] A. Watkins and L. Boggess, "A new classifier based on resource limitedartificial immune systems," in Proc. Congr. Evol. Comput., May 2002,pp. 1546–1551.

R.Sridevi is Associate Professor and Head with the Department of Information and Technology at Shri Angalamman College of Engineering and Technology. She is currently pursuing her doctorate in PeriyarManiammai University, Tanjore, India. Her area of interest includes data mining and networking.

Dr.G.Jagajothi is currently working as Head of the Department of Information Technology in PeriyarManiammai University, India. He has completed M.Tech in Laser and Electro Optical Engineering in CEG, Anna University and Ph.D in ECENIT, Trichy. He has been working as faculty of Electronics and Communication since 1996.