Ved Prakash Sharma et al. International Journal of Recent Research Aspects ISSN: 2349-7688, Vol. 5, Issue 1, March 2018, pp. 90-99

Immunity based Intrusion Detection System using Probabilistic Dendritic Cell Algorithm

Ved Prakash Sharma[*1], Rajdev Tiwari[#2]
[*]Department of IT, Northern India Engineering College, New Delhi, India
[1]mtech.ved@gmail.com
[#]Department of CSE, GNIOT, Greater Noida, India
[2]rajdevtiwari@yahoo.com

***Abstract* - Malicious intruders on networks have been the major concerns since last one decade or more, to gain the focus of many researchers. Inspired from human immune system, Dendritic Cell Algorithm has already been implemented for intrusion detection systems. Dendritic cells are the sole of human immune system which are responsible for combining signals in the tissue and reporting to the immune system of any variation in the signal concentration. In this paper an intrusion detection system is proposed and implemented for KDD data set using probabilistic dendritic cell algorithm. Besides taking statistically derived input signal concentration, probabilistic measure based on Shanon's entropy are taken as input signal concentration to compute the output cytokines. Proposed method has shown significant improvement as compared to other existing techniques used in intrusion detection system.**

Keywords - Intrusion Detection System, Immune System, Dendritic cell Algorithm, Anomaly detection, Shanon's entropy etc.

I. Introduction

An intrusion detection system (IDS) is a system to keep eyes on the network traffic and identifies suspicious patterns, called as intruders. It then generates the security alerts accordingly. Earlier, this task was performed manually where system administrator used to sit for hours with the log files to identify the potential threats to the computer system. Machine learning techniques have made this task easy by paving the way for automated IDS. Several machine learning techniques has been applied to anomaly detection, including neural networks and statistical learning algorithms. IDS can be categorized in several ways; one such categorization is signature based IDS and anomaly based IDS. A signature-based IDS monitors traffic on the network and compares them with already known malicious threats. Signatures (i.e. characteristics) of such known threats are stored in a database for reference. In anomaly-based IDS, network traffic is monitored and compared with an established profile of the traffic. The profile indicates whether network traffic is normal or not. It generally monitors that how much bandwidth is being used, what protocols are used, what ports and devices generally connect to each other and generates an alert when traffic is found anomalous (i.e. deviated significantly from the established profile) [1].

A huge amount of research work has been contributed to the area of anomaly detection because of its large application domains such as system health management, intrusion detection, health-care, bio-informatics, fraud detection, and mechanical fault detection [2]. Anomaly detection systems based on evolutionary and statistical techniques have large rate of false alarms (false positives), generally; but on the other hand they are capable of identifying even novel attacks. Artificial Immune Systems (AIS), inspired from human immune system, have also been applied to anomaly detection. In [3], Denning provided a model with the expectation that model will provide a sound basis for developing a powerful real-time intrusion detection capable of detecting a wide range of intrusions related to attempted break-ins, masquerading (successful break-ins), system penetrations, Trojan horses, viruses, leakage and other abuses by legitimate users, and certain covert channels. [4] Presents the results of a research effort that investigated the application of an adaptive neural network in the detection of network attacks which demonstrate the potential for a powerful new analysis component of a complete intrusion detection system that would be capable of identifying priori and a priori denial of service attack patterns.

A prototype presented in [5] demonstrates that IDS can be viewed as multiple function entity and can be encapsulated as autonomous agents. It is further demonstrated in this paper that genetic programming can be used as a learning paradigm to train our autonomous agents to detect potentially intrusive behaviors. Paper [6] proposes a novelty detection method inspired from human immune system. This method notices changes in normal behavior without requiring prior knowledge of the changes for which it is looking.

An artificial immune system framework (ARTIS) is described in [7], which incorporates many characteristics of natural immune system, including diversity, distributed computing, error tolerance, dynamic learning and adaptation and self-monitoring. Traditional anomaly detection techniques analyze each data instance (as a uni-variate or multivariate record) independently and ignore the sequential aspect of the data. But it is noticed that anomalies in sequences can be detected only by analyzing data instances together as a sequence, and hence cannot

be detected by traditional anomaly techniques [8]. In [9] a technique inspired from negative selection mechanism of the immune system is applied to detect foreign patterns in the complement (nonself) space. In [10], Hofmeyr developed an artificial immune system based on 'negative selection': detectors forming the normal profile are deleted if they match a string denoting normal behavior.

At the time, it was perceived to function in a similar way to the selection of T-lymphocyte cells in the thymus. Problems with negative selection were highlighted in [11] and more recently in [12]. [13] Proposes that the negative selection algorithm could not work because it was based on a simplified version of the immunological self-nonself theory. This theory has been challenged within immunology itself and an alternative theory has been proposed - the Danger Theory [14]. The Danger Theory states that the immune system does not discriminate on the basis of self or nonself, but on the balance between the concentration of danger and 'safe' signals within the tissue of the body. Idea of danger theory is applied in this paper to detect the intrusions.

In [15], dendritic cell inspired algorithm on two datasets is demonstrated and advocated with promising results that DCA plus libtissue framework can be used for the purpose of anomaly detection under real-time conditions. An approach in [16] is applied to detect anomalous activity in the network, using detectors generated by the genetic algorithm. The Minkowski distance function is tested against the Euclidean distance for the detection process. It is shown that Minkowski distance gives better results with 81.74% overall average detection rate than the Euclidean distance which gives 77.44% detection rate.

This presented paper focuses on the development of IDS, built on the foundation of immune inspired system. Dendritic cells are the sole of human immune system which are responsible for combining signals in the tissue and reporting to the immune system of any variation in the signal concentration. The signal concentration is determined using a weighted function with fixed suggested weight values obtained from empirical data based on immunologists' wet lab results (Dr Julie McLeod, Dr Rachel Harry and Charlotte Williams - University of the West of England). In this paper an intrusion detection system is proposed and implemented for KDD data set using probabilistic dendritic cell algorithm besides taking statistically derived initial concentration, probabilistic initial concentrations based on Shanon's entropy are taken into account to compute the output signal concentration. Rest of the paper is organized as follows: In section II, brief insight of immune system is presented. This section also talks about the integration of IDS and AIS. In section III, simulation design along with the implementation details are discussed. Section IV presents the results obtained and concluding remarks have been put in the section V followed by the references.

II. Immune System

Immune system is a body-wide network of cells, tissues, and organs that has evolved to defend us against foreign invasions. Earlier theory of human immune system was based on discrimination between antigens (proteins) belonging to 'self' versus antigens belonging to infectious agents called pathogens - 'nonself' [15]. At the heart of the immune response is the ability to distinguish between self and non-self. Every cell in our body carries the same set of distinctive surface proteins that distinguish us as self. Any non-self substance capable of triggering an immune response is known as an antigen. The organs of our immune system are positioned throughout our body. They are called lymphoid organs because they are home to lymphocytes--the white blood cells that are key operatives of the immune system.

Within these organs, the lymphocytes grow, develop, and are deployed. The thymus is an organ that lies behind the breastbone; lymphocytes known as T lymphocytes, or just T cells, mature there [17]. T cells contribute to your immune defenses in two major ways. Some help regulate the complex workings of the overall immune response, while others are cytotoxic and directly contact infected cells and destroy them. Chief among the regulatory T cells are helper T cells. They are needed to activate many immune cells, including B cells and other T cells.

Recently, several questions have been raised regarding the validity of this model. Definition of 'self' actually varies throughout the lifetime of an individual, e.g. a pregnant woman's immune system does not react against her unborn fetus despite consisting of 'nonself' proteins [15]. A modification to the self-nonself theory was proposed in [18], by the name 'infectious non-self' model. This states that an antigen must be associated with a PAMP (pathogen associated molecular patterns) in order to trigger a response, as recognized by the first line of defense (i.e. innate immune system) against invading organisms. Though this model could provide the logic for adding stimulatory adjuvant to vaccines yet it could still not answer pertinent questions relating to autoimmunity. Danger Theory [14] provides an alternative view of the activation of the immune system. According to danger theory the immune system detects the presence of danger signals, released as a result of necrotic cell death within the host tissue.

A. Danger Theory

The Danger Theory proposes that the immune system is sensitive to changes in the danger signal concentration in the tissue. Conversely, when then tissue is healthy, cells die in a controlled manner, known as apoptosis. Immunosuppressive molecules (safe signals) are released as an indicator of normality in the tissue. In essence, the Danger Theory consists of active suppression while the tissue is healthy (apoptosis), combined with rapid

activation on receipt of necrotic danger signals. Necrosis is the result of cellular damage and stress caused by pathogenic infection or exposure to extreme conditions. The metabolites of internal cell components are thought to form the danger signals and are released into the surround buffer fluid. The cell membrane loses its integrity, releasing its contents (e.g. DNA, mitochondria) into the surrounding tissue fluid [19]. This property is abstracted to form artificial tissue, as conceptualized in a software framework in [20].

B. Dendritic Cells

Dendritic cells (DCs) are white blood cells that reside in tissue in immature state. DCs function is to collect antigen from pathogens, experience danger signals from necrosing cells and 'safe' signals from apoptotic cells. Maturation of DCs occurs in response to the receipt of these signals. Once DC is reached in mature state, it stops further collection of antigen; expresses the costimulatory molecules (CSM) and cytokines; migrate from the tissue to lymph node; and present antigen to T-lymphocytes. If there is a greater concentration of danger signals in the tissue at the time of antigen collection, the DC will become fully mature (mDC), and will express mDC cytokines; if the DC is exposed to 'safe' signals; it becomes semi-mature and expresses smDC cytokines [21].

C. The DC Algorithm

Some functions of DCs are abstracted to form an algorithm. Key functions used in dendritic cell algorithm (DCA) are as bellow;

- Immature DCs collect antigens from multiple sources and are exposed to signals in the host tissue.
- DCs are capable to combine signals from multiple sources to generate different output concentrations of CSM, smDC cytokines and mDC cytokines.
- Capable of increasing CSMs that lead to migration of DCs to the lymph node.
- Exposure to signals causes the maturation of DCs into either mature and semi mature states.

A simple interpretation of the input signals has been derived. There are four signals in the model given in [15], each from a different source and producing different output cytokines:

- PAMPs (P) are based on pre-defined signatures. Exposure to PAMPS causes an increase in mDC cytokines. PAMPs are suppressed by safe signals.
- Danger signals (D) cause an increase in mDC cytokines. Danger Signals can also be suppressed by safe signals. Danger signals have a lower potency than PAMPs.
- Safe signals (S) cause an increase in smDC cytokines and have a suppressive effect on both PAMPS and danger signals.
- Inflammatory cytokines (IC) amplify the effects of the other three signals, but are not sufficient to cause any effect on DCs when used in isolation.

Their data and method of processing is very different from other AIS, which rely on negative selection [23] or on pattern matching of antigen to drive their systems [24]. In their algorithm, representation of the antigen can be a string of either integers or characters. Signals are represented as real-valued numbers, proportional to values derived from the context information of the dataset in use. The signal values are combined using a weighted function with suggested values of the weights derived from empirical data based on immunologists' wet lab results (Dr Julie McLeod, Dr Rachel Harry and Charlotte Williams - University of the West of England). The function itself is a weighted sum of PAMP, danger and safe signal concentration values, multiplied by a value for inflammation (in the range of 0 and 2).

The resulting value is then normalized through division by the sum of the weights. The function is used three times to calculate the output cytokines of CSMs, mDC and smDC cytokines. Transition to the mature state depends on the CSM value. Each cell is assigned an individual migration CSM threshold value, which can vary between cells. When a cell's CSM value exceeds the migration threshold, the status of the cell changes and migration from the tissue is initiated. In figure1 migration of DCs are shown diagrammatically.

The DCA is a population based algorithm; where population size (i.e. number of DCs) is defined by the user create a pool. Each DC in the pool is exposed to current signal values and selects a slot in the antigen store. If an antigen is present in the antigen store, the DC collects the antigen and ingests it in the DC internal antigen storage. Each DC has the opportunity to sample multiple antigens. Antigen collection is done repeatedly and each DC re-calculates its output signal concentration. Migration is initiated by the removal of a DC from the pool and occurs when the cell's internal CSM value exceeds the DC's migration threshold.

At this point, the output cytokines of each DC are measured. Antigen presented by cells predominantly expressing mature cytokines is labeled 'mature context antigen', where as antigen from cells expressing predominantly semi-mature cytokines is labeled as 'semimature'. Each presented antigen's context is recorded and finally a mean antigen context value (between 0 and 1) is derived. In our model, Shanon entropy is taken as a probabilistic measure for input signal concentration combined with conventional DCA as discussed above to investigate the performance of IDS.

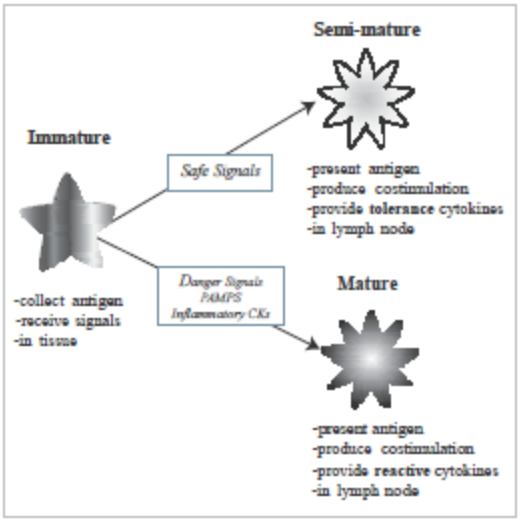

Figure1: Behavior of DCs

III. Simulation design and implementation details

A. Shanon's Entropy

It is an information-theoretic measure of uncertainty, variability or complexity of a collected data set. For a system X with a finite set of M possible states,$\{x_1, x_2, x_3, \ldots \ldots x_M\}$ the Shanon entropy of X is defined as;

$H(X) = -\sum_{i=1}^{M} P(x_i) log\big(P(x_i)\big)$

where, $P(x_i)$ is the probability that the system X is in state x_i. Shanon entropy typically interpreted as average information content of the data source. Entropy value is smaller when data distribution is skewed that is when data exhibits a pattern. Entropy is larger when data is symmetric, that is, when data exhibits randomness [25].

B. Probabilistic DCA

Probabilistic dendritic cell algorithm as given below is a modified form of DCA given in [15 & 22]. An information-theoretic measure, Shanon entropy is computed for the data items (also referred as antigens in immune system terminology) categorized into three types of signals, named PAMP, Danger signal & Safe signal and is denoted by $C_p, C_d \wedge C_s$ respectively by using eqns 1-3. Corresponding output concentrations for three input signals are computed using eqns 4-6. On the basis of output computed cytokines, migration of dendritic cell from immature state to mature state or to semi-mature state takes place. Finally, count of mature and semi-mature antigens decides whether the antigen is malicious or normal. Below mentioned algorithm is implemented using Java.

P_DCA()
{

1. *Define an empty DC pool with no cell defined internally.*
2. *L=Length(AvailableDataItems)*
3. *For i=1 to L*
4. *{*
5. *Select CurData=AvailableDataItems(i)*
6. *compute_mean(AvailableDataItems(i))*
7. *Classify the classes into PAMP, Safe and Danger on the basis of deviation in the mean from normal class.*
8. *Compute initial concentration by Shanon's formula for PAMP, safe and Danger signals;* $C_p, C_s \wedge C_d$.
9. *Compute the signal concentrations for CSM, mDC and smDC.*
10. *If (OutputCytokines>SafeLimit)*
11. *{*
12. *Migrate Signal to Safe Signal Dataset*
13. *}*
14. *Else If(OutputCytokines<SafeLimit And OutputCytokines<PAMPLimit)*
15. *{*
16. *Migrate Signal to PAMP Signal Dataset*

17. *}*
18. *Else*
19. *{*
20. *Migrate Signal to Danger Signal Dataset*
21. *}*
22. *Obtain MeanDangerSignal from Derived OutputCytokines*
23. *If semi>mature*
24. *{*
25. *Set AntigenContext=Semi*
26. *}*
27. *Else*
28. *{*
29. *Set AntigenContext=Mature*
30. *}*
31. *For i=1 to length(AntigenContext)*
32. *{*
33. *Identify the number of appearances of antigen as semi or mature*
34. *If semi>mature*
35. *{*
36. *Set Antigen=normal*
37. *}*
38. *Else*
39. *{*
40. *Set Antigen=malicious*
41. *}*
42. *}*

C. Data set

KDD data set is considered for the purpose of simulation in this paper. It has total 110570 records with 39 attribute out of which last attribute is the class label. There are 23 classes in which all data items are classified including the class 'normal'. In table 1 & 2, means of some of the attributes class-wise is shown and is also depicted in the figure 2 & 3 respectively. This is done for the purpose of categorizing data items into three signal categories, named PAMP, Safe and Danger. Normal classed data items are mapped to safe signals. Data items with large deviated mean values as compared to normal classed data are mapped to PAMP and rest to danger signals.

Table1: Attribute means for some classes.

	warezmaster	warezclient	teardrop	spy	smurf	satan	rootkit	portsweep	pod	phf	perl	normal
land	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
wrong_fragment	0.0000	0.0000	2.9823	0.0000	0.0000	0.0000	0.0000	0.0000	0.9821	0.0000	0.0000	0.0000
urgent	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1000	0.0000	0.0000	0.0000	0.0000	0.0001
num_failed_login	0.0000	0.0000	0.0000	0.0000	0.0000	0.0012	0.1000	0.0000	0.0000	0.0000	0.0000	0.0013
num_compromised	0.0000	0.0000	0.0000	0.0000	0.0000	0.0006	1.3000	0.0000	0.0000	0.0000	0.0000	0.4969
root_shell	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2000	0.0000	0.0000	1.0000	1.0000	0.0020
su_attempted	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0019
num_root	0.0000	0.0000	0.0000	0.0000	0.0000	0.0006	2.7000	0.0000	0.0000	0.0000	2.0000	0.5494
num_shells	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0006

num_access_files	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0074
num_outbound_cmds	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
is_host_login	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
serroe_rate	0.0000	0.0026	0.0735	0.0000	0.0000	0.0490	0.0000	0.0350	0.0000	0.0000	0.0000	0.0134
srv_serror_rate	0.0000	0.0027	0.0000	0.0000	0.0000	0.0274	0.0000	0.0364	0.0000	0.0000	0.0000	0.0119
rerror_rate	0.0000	0.0032	0.0016	0.0000	0.0000	0.5244	0.0000	0.9449	0.0000	0.0000	0.0000	0.0441
srv_rerror_rate	0.0000	0.0026	0.0000	0.0000	0.0000	0.5438	0.0000	0.9474	0.0000	0.1667	0.0000	0.0446
diff_srv_rate	0.0000	0.0085	0.0099	0.0000	0.0000	0.6619	0.0000	0.1964	0.0000	0.0000	0.0000	0.0291
srv_diff_host_rate	0.0000	0.0138	0.0000	0.0000	0.0000	0.0023	0.0000	0.0019	0.2845	0.3333	0.0000	0.1258
dst_host_srv_diff_host_rate	0.0000	0.0944	0.0000	0.0000	0.0000	0.0009	0.0100	0.0013	0.1717	0.0000	0.0000	0.0260
dst_host_srv_serror_rate	0.0000	0.0041	0.0000	0.1550	0.0000	0.0269	0.0250	0.0359	0.0000	0.0000	0.0000	0.0061
dst_host_srv_rerror_rate	0.0000	0.0004	0.0000	0.0000	0.0000	0.5434	0.0250	0.9397	0.0000	0.0000	0.0000	0.0448

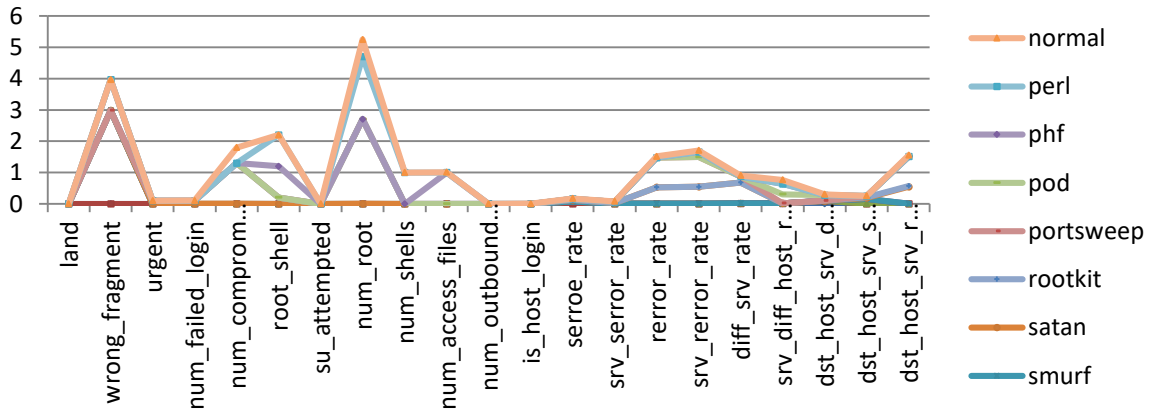

Figure2: Deviation of means for some classes as compared to normal class

Table2: Attribute means for some other classes

	nmap	neptune	multihop	loadmodule	land	ipsweep	imap	guess_passwd	ftp_write	buffer_overflow	back	normal
land	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
wrong_fragment	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
urgent	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.4286	0.0000	0.0000	0.0001
num_failed_login	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0714	0.0000	0.0000	0.0000	0.0013
num_compromised	0.0000	0.0000	8.5714	0.7143	0.0000	0.0013	1.4545	0.0000	0.1429	1.5769	0.9224	0.4969
root_shell	0.0000	0.0000	0.2857	0.4286	0.0000	0.0000	0.0000	0.0000	0.0000	0.6923	0.0000	0.0020

su_attempted	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0019
num_root	0.0000	0.0000	13.2857	0.4286	0.0000	0.0013	1.4545	0.0000	0.2857	0.0385	0.0000	0.5494
num_shells	0.0000	0.0000	0.4286	0.5714	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0006
num_access_files	0.0000	0.0000	0.2857	0.1429	0.0000	0.0000	0.0000	0.0000	0.4286	0.0000	0.0000	0.0074
num_outbound_cmds	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
is_host_login	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
serroe_rate	0.1700	0.8318	0.0000	0.0000	0.9553	0.0003	0.6673	0.0476	0.0000	0.0381	0.0064	0.0134
srv_serror_rate	0.1716	0.8312	0.0000	0.0000	1.0000	0.0000	0.5773	0.0476	0.0000	0.0000	0.0070	0.0119
rerror_rate	0.0000	0.1678	0.0000	0.0000	0.0447	0.1184	0.0000	0.9048	0.0000	0.0196	0.0773	0.0441
srv_rerror_rate	0.0000	0.1675	0.0000	0.0000	0.0000	0.1183	0.0255	0.9048	0.0000	0.0385	0.1357	0.0446
diff_srv_rate	0.0101	0.0725	0.0000	0.1671	0.0720	0.0022	0.0000	0.0000	0.0000	0.0408	0.0040	0.0291
srv_diff_host_rate	0.6430	0.0014	0.1429	0.0000	0.8000	0.7020	0.2345	0.0000	0.1429	0.0000	0.1224	0.1258
dst_host_srv_diff_host_rate	0.1657	0.0007	0.0000	0.1714	0.5193	0.5375	0.0000	0.0238	0.0771	0.0865	0.0000	0.0260
dst_host_srv_serror_rate	0.1704	0.8291	0.0000	0.0000	0.6380	0.0000	0.5791	0.1107	0.0000	0.0000	0.0021	0.0061
dst_host_srv_rerror_rate	0.0000	0.1673	0.0000	0.0143	0.0000	0.1134	0.0000	0.8650	0.0000	0.0246	0.0621	0.0448

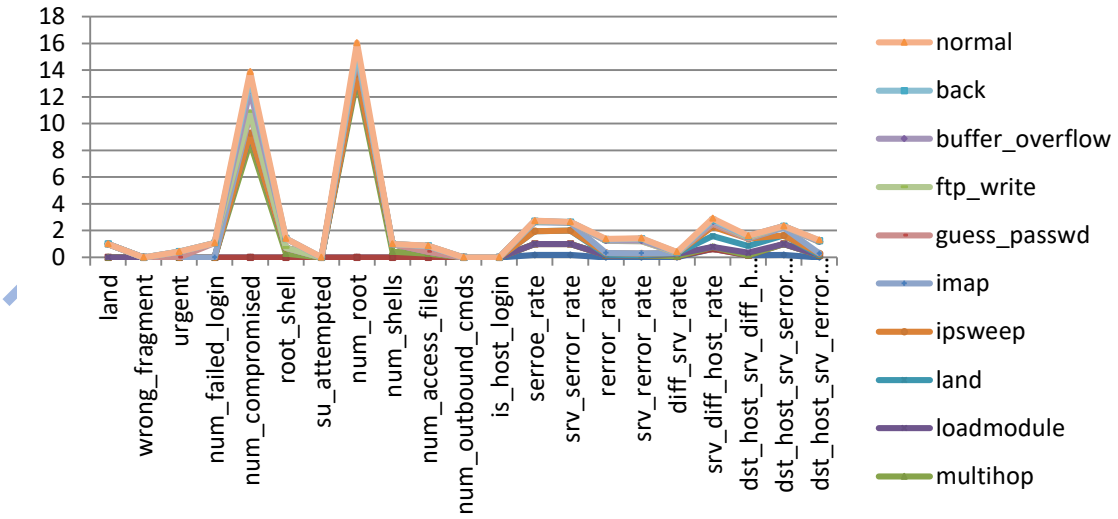

Figure3: Deviation of means for some classes as compared to normal class.

D. Entropy as Input Signal

Shanon entropies for all three signals are computed by using eqns 1-3. Entropy of whole data set computed is shown the table 3.

$$p_i(X)log\big(p_i(X)\big)(1)$$

$$C_p = -\sum_{i=1}^{l}$$

where X is collection of l data items categorized as PAMP.

$$p_i(Y)log\left(p_i(Y)\right)(2)$$

$$C_s = -\sum_{i=1}^{m}$$

where Y is collection of m data items categorized as safe.

$$p_i(Z)log\left(p_i(Z)\right)(3)$$

$$C_d = -\sum_{i=1}^{n}$$

where Z is collection of n data items categorized as dander.

Table3: Entropy of the system.

Data Item Count	P(X)	log(P(X))	P(X)*log(P(X))
2.0000	0.0000	-4.7426	-0.0001
839.0000	0.0076	-2.1199	-0.0161
26.0000	0.0002	-3.6287	-0.0009
7.0000	0.0001	-4.1985	-0.0003
42.0000	0.0004	-3.4204	-0.0013
11.0000	0.0001	-4.0022	-0.0004
3171.0000	0.0287	-1.5424	-0.0442
15.0000	0.0001	-3.8675	-0.0005
7.0000	0.0001	-4.1985	-0.0003
7.0000	0.0001	-4.1985	-0.0003
36221.0000	0.3276	-0.4847	-0.1588
1294.0000	0.0117	-1.9317	-0.0226
59057.0000	0.5341	-0.2724	-0.1455
3.0000	0.0000	-4.5665	-0.0001
3.0000	0.0000	-4.5665	-0.0001
168.0000	0.0015	-2.8183	-0.0043
2570.0000	0.0232	-1.6337	-0.0380
10.0000	0.0001	-4.0436	-0.0004
3203.0000	0.0290	-1.5381	-0.0446
2324.0000	0.0210	-1.6774	-0.0353
789.0000	0.0071	-2.1466	-0.0153
18.0000	0.0002	-3.7884	-0.0006
783.0000	0.0071	-2.1499	-0.0152
Entropy of the System,H(X)=			0.5450

E. Signal Concentration and Maturation Table: The signal values are combined using a weighted function (Equation 4-6) with suggested values of the weights derived from empirical data based on immunologists' wet lab results (Dr Julie McLeod, Dr Rachel Harry and Charlotte Williams - University of the West of England). These empirically derived weights are shown in table 4.

$$C_{CSM} = \frac{\left(w_{p,csm} \times C_p\right) + \left(w_{s,csm} \times C_s\right) + \left(w_{d,csm} \times C_d\right)}{\left(w_{p,csm} + w_{s,csm} + w_{d,csm}\right)}(4)$$

$$C_{mDC} = \frac{\left(w_{p,mDC} \times C_p\right) + \left(w_{s,mDC} \times C_s\right) + \left(w_{d,mDC} \times C_d\right)}{\left(w_{p,mDC} + w_{s,mDC} + w_{d,mDC}\right)}(5)$$

$$C_{smDC} = \frac{(w_{p,smDC} \times C_p) + (w_{s,smDC} \times C_s) + (w_{d,smDC} \times C_d)}{(w_{p,smDC} + w_{s,smDC} + w_{d,smDC})} 6$$

Table 4: Empirically derived weights.

W	csm	semi	mat
PAMPs(P)	2	0	2
Danger Signals(D)	1	0	1
Safe Signals(S)	2	3	-3

IV. Result Discussion

As shown in figure 4 & 5, classification accuracy of the proposed P-DCA based IDS is higher most as compared to some well known techniques earlier used in IDS like; support vector machine (SVM), Multi layer perceptron (MLP), Random tree and Naïve Bays. P-DCA based IDS is simulated and tested for KDD+ and KDD21 the two versions of KDD data set and the superiority of proposed algorithm is hence validated. Proposed P-DCA based model gives 92.58% accuracy for KDD+ data set which is highest amongst the considered technologies. This result is validated by the result for KDD21 data set where 61.05% accuracy is achieved. No other technique considered for comparison could provide accuracy even upto 60%.

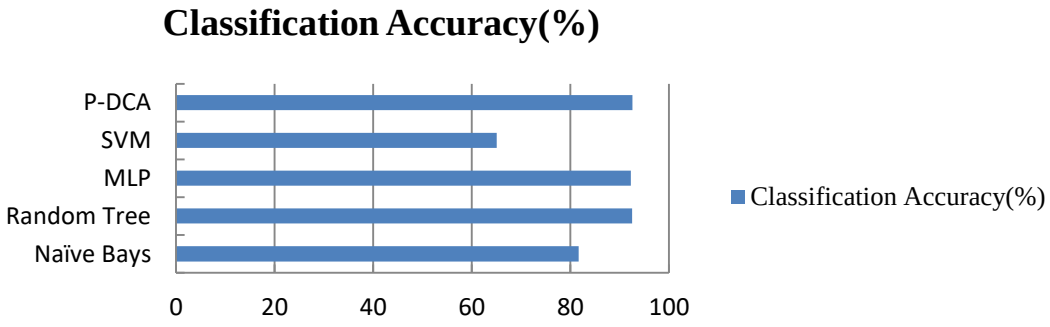

Figure 4: For data set KDD+

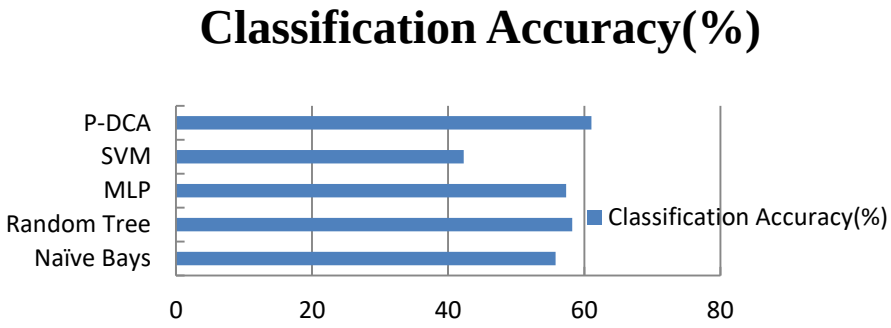

Figure5: For data set KDD21

V. Conclusions

Probabilistic DCA based intrusion detection system proposed and implemented in this paper for two versions of KDD data set. In this paper functionalities of dendritic cells are abstracted to model the P-DCA where input signal concentration is actually the Shanon entropy for respective categories (i.e. PAMP, Safe & Danger) of signals. Results advocate that P-DCA based method implemented and validated in this paper for intrusion detection improves the detection accuracy significantly as compared to the earlier techniques being used for the same. Proposed IDS gives 92.58% detection accuracy where as SVM, MLP, Random tree and Naïve Bays based IDS give 65.01%, 92.26%, 92.53% and 81.66% respectively when implemented on KDD+ data set. For KDD21 data set detection accuracy for proposed method was found to be 61.56% where as for SVM, MLM, Random Tree and Naïve Bays based methods, it was 42.29%, 57.34%, 58.21% and 55.77% respectively.

References

[1] Mr. Vedprakash sharma, Dr. Rajdev Tiwariki. Dendritic cell algo and Dempster belief theory based approach for IDS. Journal of computer engg.(IOSR) e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 2, Ver. IX (Mar-Apr. 2014), PP 99-103
[2] V. chandola, A. Banerjee, V. Kumar, Anomaly detection: A survey, ACM computing sureys, 2009
[3] D. Denning, "An intrusion-detection model," *IEEE Trans. Software Eng.*, vol. 13, pp. 222–232, Feb. 1987.
[4] J. Cannnady, "Next generation intrusion detection: Autonomous reinforcement learning of network attacks," in *Proc. 23rd Nat. Information Systems Security Conf.*, pp. 1–12, Oct. 2000.
[5] M. Crosbie and E. H. Spafford, "Applying genetic programming to intrusion detection," in*Working Notes for the AAAI Symposium on Genetic Programming*, E. V. Siegel and J. R. Koza, Eds. Cambridge, MA: MIT Press, pp. 1–8, 1995.
[6] D. Dasgupta and S. Forrest, "Novelty detection in time series data using ideas from immunology," in *Proc. Int. Conf. Intelligent Systems*, pp. 87–92, June 1996.
[7] S. A. Hofmeyr and S. Forrest, "Architecture for an artificial immune system," *Evol. Comput.*, vol. 8, no. 4, pp. 443–473, 2000.
[8] Varun Chandola, Anomaly Detection for Symbolic Sequences and Time Series Data, PhD. Dissertation. Computer Science Department, University of Minnesota, http:// purl.umn.edu/56597. 2009
[9] Fabio A. Gonzalez and Dipankar Dasgupta, An Immunity-based Technique to Characterize Intrusions in Computer Network, IEEE Transactions on Evolutionary Computation, Vol. 6(3), pp. 281-291, 2002.
[10] J Kim and P J Bentley. Towards an artificial immune system for network intrusion detection: An investigation of clonal selection with a negative selection operator. In Proceeding of the Congress on Evolutionary Computation (CEC-2001), Seoul, Korea, pages 1244–1252, 2001.
[11] S Hofmeyr. An immunological model of distributed detection and its application to computer security. PhD thesis, University of New Mexico, 1999.
[12] T Stibor, P Mohr, J Timmis, and C Eckert. Is negative selection appropriate for anomaly detection? In Proceedings of Genetic and Evolutionary Computation Conference (GECCO) Washington DC. USA. pages 321–328, 2005.
[13] U Aickelin, P Bentley, S Cayzer, J Kim, and J McLeod. Danger theory: The link between ais and ids. In Proc. of the Second International Conference on Artificial Immune Systems (ICARIS-03), pages 147–155, 2003.
[14] P Matzinger. Tolerance, danger and the extended family. Annual Reviews in Immunology, 12:991–1045, 1994.
[15] Julie Greensmith, Jamie Twycross and Uwe Aickelin. Dendritic cells for Anomaly detection, IEEE congress on Evolutionary Computation, Vancouver, BC, Canada, 16-21, 2006.
[16] Amira sayed A.Aziz, M. a. Salman, A. ella Hassanien, Sanaa El-Ola Hanafi. Artificial Immune System Inspired Intrusion Detection System Using Genetic algorithm, Informatica 36,347-357,2012.
[17] http://cancer.gov/cancertopics/understandingcancer
[18] Charles A Janeway. Approaching the asymptote? Evolution and revolution in immunology. Cold Spring Harb Symp Quant Biol, 1:1–13, 1989
[19] P Matzinger. The danger model: A renewed sense of self. Science, 296:301–304, 2002.
[20] J Twycross and U Aickelin. libtissue - implementing innate immunity. In To Appear in the Proceedings of the IEEE Congress on Evolutionary Computation, Vancouver, Canada., 2006.
[21] T R Mosmann and AM Livingstone. Dendritic cells: the immune information management experts. Nature Immunology, 5(6):564–566, 2004.
[22] U Aickelin, J Greensmith, and J Twycross. Immune system approaches to intrusion detection - a review. In Proc. of the Third International Conference on Artificial Immune Systems (ICARIS-04), pages 316–329, 2004.
[23] T Stibor, P Mohr, J Timmis, and C Eckert. Is negative selection appropriate for anomaly detection? In Proceedings of Genetic and Evolutionary Computation Conference (GECCO) Washington DC. USA. pages 321–328, 2005.

