ISSN:0975-9646

Analysis of Human Immune System Inspired Intrusion Detection System

R.Sridevi[#1] Dr. Rajan Chattamvelli[#2] Dr.E.Kannan[*3]

[#1] [#2] *Periyar Maniammai University*
Thanjavur 613403 INDIA '

[*3] *Registrar*
VEL TECH Dr.RR & Dr.SR TECHNICAL UNIVERSITY
Chennai 600 062,

Abstract — Artificial Immune Systems (AIS) are algorithms inspired by the human immune system. The human immune system is a robust, decentralized, error tolerant and adaptive system. Such properties are highly desirable for the development of novel computer systems. Unlike some other bio-inspired techniques, such as genetic algorithms and neural networks, the field of AIS encompasses a spectrum of algorithms to implement different functions. In this paper we investigate CLONALG for network intrusion classification. The Clonal Selection Algorithm (CLONALG) is inspired by the clonal selection theory of acquired immunity, which has shown success on broad range of engineering problem domains.

***Keywords*— Network intrusion detection, Artificial Immune System, CLONALG, KDD dataset.**

I. INTRODUCTION

With information explicitly available in the Internet than ever before, the risk associated with network attacks by hackers are increasing. The emerging and existing technology of firewalls or authentication systems is no longer enough for the security of the existing system. The conventional intrusion detection systems are built by incorporating information available from the system which provides a comprehensive definition of normal activities called Self, which the system uses to differentiate unusual activities, categorized as Nonself. Normally, IDS are based on the principle that an intruder's behaviour will be visibly different from that of a legitimate user and that the many unauthorized actions are distinguishable. Thus it is possible to determine whether there is an attack occurring in the system by means of collecting and analyzing operating system activity data and network activity data [1].The classification of system activities falls into four categories:

- IDS monitors and analyzes user's activity and system activity
- Assess the integrity of the critical system and data files
- To recognize activity patterns reflecting known attacks.
- To respond automatically on detecting activity, and
- To report the outcome of the detection process [4]

Data mining, at its outset, is a pattern finding tool. Data miners are experts at using specialized software to find regularities (and irregularities) in large data sets. Applications of Data mining in computer security concentrate heavily on the use of data mining in the area of intrusion detection. The volume of data dealt in both network and host activity is so large, that data mining is ideal for mining critical activity. An ultimate application of intrusion detection will be to gather sufficient "normal" and "abnormal" audit data for a user or a program, and then apply a classification algorithm to learn a classifier that will analyse (future) audit data to decide whether it belongs to the normal class or the abnormal class [5].

Here are a few specific things that data mining might contribute to intrusion detection.

- Remove normal activity from alarm data to allow analysts to focus on real attacks
- Identify false alarm generators and "bad" sensor signatures
- Find anomalous activity that uncovers a real attack
- Identify long, ongoing patterns (different IP address, same activity)

To accomplish these tasks, data miners use one or more of the following techniques:

- Data summarization with statistics, including finding outliers
- Visualization: presenting a graphical summary of the data
- Clustering of the data into natural categories
- Association rule discovery to define normal activity and enabling the discovery of anomalies

- Classification: predicting the category to which a particular record belongs.

Data mining and intrusion detection, the two methods are capable of working together efficiently to provide network security. Data mining can improve a network intrusion detection system by detecting network data differences. Data mining provides an extra level of intrusion detection by identifying the boundaries for usual network activity so it can distinguish common activities from uncommon intrusion. Different data mining methods used in IDS are:

1. Code Variants: Data mining process of scans for abnormal activity through code variants instead of unique signatures.
2. Data Reduction: In order to extract the relevant data and to avoid data overloading, data reduction techniques are applied and the obtained data is utilized for identification and analysis.
3. Filter out Valid Network Activity: Data mining helps intrusion detection to work efficiently by identifying valid network activity; so it can filter it out to make detection of abnormal activity in data easier.
4. Attacks without Signatures: If network activity contains a specific profile and rules of protocol, an abnormality is easily detected and monitoring of individual hosts, entire networks, specific users, and overall traffic patterns on the network at specific times is easily accomplished. Data mining is more efficient in detecting abnormalities that do not contain signatures [6].

Classification Techniques

Classification is used to assign examples to predefined categories. Machine learning software performs this task by extracting or learning discrimination rules from examples of correctly classified data. Classification models are built using different kinds of algorithms. Steven et.al., classifies classification algorithms into three types: [7]

- Extensions to linear discrimination (e.g., multi- layer perceptron, logistic discrimination),
- Decision tree and rule-based methods (e.g. C4.5, AQ, CART), and
- Density estimators (Naïve Bayes, k-nearest neighbor, LVQ) [8].

Classifiers are the tools that partition the given data sets into different categories on the basis of specified features [9]. Classifier selection is one of the important research challenge faced in building efficient IDS. Nowadays, many data mining algorithms have become very popular for its effective classification of intrusion detection datasets such as decision tree, naïve Bayesian classifier, neural network, genetic algorithm, and support vector machine etc. However, the classification accuracy of most existing data mining algorithms needs to be improved, because it is very difficult to detect several novel attacks. Anomaly network intrusion detection models are now used to detect new attacks but the false positives are found to be very high. The intrusion detection model efficiency depends on its detection rates (DR) and false positives (FP). DR is defined as the number of intrusion instances detected correctly by the system divided by the total number of the intrusion instances present in the given dataset. FP is an alarm, which indicates attack when it is not really an attack. The aim of an intrusion detection model is to maximize the DR and minimize the FP. Classifier construction for IDS is a technical challenge in the field of data mining. Dewan Md. Farid et al., proposed a new classification model which gives higher accuracy when compared with existing models. The proposed model had 99.65% accuracy when tested on correctly classified instances having 41 attributes, when compared to ID3 Algorithm 99.63% and Naïve Baysian 99.27% [10].

II. IMMUNOLOGY

Immunologists believe that 'immunity is to identify the I (Self) and nonself (Nonself), and eliminate nonself as a physiological response, to ensure the integrity of the body'. Human's natural immune system includes the skin, physiological conditions, congenital immune system and adaptive immune system. The skin is the first line of defence to prevent disease; physiological conditions is the second line of defence ; once pathogens enter the body, then would face the third and fourth lines of defence.

The human immune system (HIS) performs many of the same functions as an intrusion detection system. Forrest et.al., first explored the idea of human immune concepts applied to computer security [11]. In order to apply these concepts, a basic understanding of how the human immune system works must be understood. The studies show that the IDS based on human immune system concepts, perform tasks similar to HIS innate and adaptive immunity. The profile of normal behaviour is generated by collecting appropriate behaviour of services represented from audit data. Nageswara Rao et al., dealt with different techniques used in the intrusion detection and suggested that different techniques like Baysian decision, Neural Networks, Fuzzy logic and Expert system can be combined for classification [12]. Zhao et al., describes several research and trend of immunity based network intrusion detection systems. Anomaly based intrusion detection attacks and the framework required to detect the attack is described. Diversity, distribution, locality, adaptability, dynamic nature are the properties obtained by using the proposed framework [13]. The primary function of the human immune system may be viewed as differentiating between things that belong in the body and that do not. The human immune system distinguishes between self and non-self antigen. The process of detecting and removing non-self involves both innate and adaptive immunity. The innate components are nonspecific and unchanging with repeated exposure to antigen. Adaptive immunity is specific and includes memory that allows the

immune system to respond more quickly the second time an antigen is encountered. John M.Hall et al., evaluated immune based system by introducing an architecture with the two systems and it shows promising result with the tested system [14].

There is a strong association between an immune system and a computer system. The immune system protects the body from pathogens and in the same way the computer security system protects the computer from malicious users. Artificial immune system is a new methodology that is increasingly attracting attention for monitoring engineered systems. In an artificial immune system (AIS), processes of the natural immune system are applied in solving real world problems.

Kaushi Ghosh et al. proposed an immune system method to diagnosis and monitor the system with

- complete fault coverage,
- Very high overall recognition rate,
- Low false positive rate,
- High true positive rate, and
- Early fault detection and diagnosis.

On comparison of performance with traditional principal component analysis (PCA) based approaches, AIS perform better [15]. The Artificial Immune Systems (AIS) are a relatively new area of research with considerable potential in help solving myriad problems. Its growth has allowed the proposal of new techniques and approaches for solving known problems. The aim of this AIS technology is to model defence mechanism characteristics and functionalities of living beings. The defence mechanism allows an organism to guard against invasion from foreign substances. The recognition of these substances is based on the key and lock analogy, in which the objective is to find antibodies that have the best immune response to the invading antigens [16].

The natural immune system stores the most effective antibodies in its genetic memory. These are used to identify antigens that have previously invaded the organism, thereby obtaining a quicker, more efficient response. New functionalities observed in the biological environment were studied for the modelling of this new immunological approach, principally the organization and clustering of similar antibodies throughout the process. It is believed that these functionalities improve the recognition capacity of artificial immune algorithms [17].

There are wide varieties of AIS applications are there. The following few are among the list:

Pattern matching
Clustering
Web data mining
Association memory
Optimization
Dynamic learning
Anomaly detection
Pattern detection [18]

The common techniques adopted by specific immunological theories, explains the function and behaviour of the mammalian adaptive immune system. There are several techniques of AIS available based on these various immunological theories. Some of them are clonal selection algorithm, Immune Network Algorithms, Dendritic Cell Algorithms [19]. This paper specifically focuses on clonal selection algorithm.

Clonal Selection Algorithms (CSAs) are a special class of Immune algorithms (IA) which are inspired by the Clonal Selection Principle [8] of the human immune system to produce effective methods for search and optimization [20].

The main features of the clonal selection theory that are used in intrusion detection are:

- Proliferation and differentiation on stimulation of cells with antigens;
- Generation of new random genetic changes, subsequently expressed as diverse antibody patterns, by a form of accelerated somatic mutation (a process called affinity maturation); and Elimination of newly differentiated lymphocytes carrying low affinity [21].

Clonal selection algorithms have taken inspiration from the antigen driven affinity maturation process of B cells and the associated hyper mutation mechanism and the idea of memory cells to retain good solutions to the problem being solved. It highlights two important features of affinity maturation in B cells that can be exploited from the computational viewpoint. The first feature is that the proliferation of B cells is proportional to the affinity of the antigen that binds it, thus the higher the affinity, the more clones that are produced. Secondly, the mutations of a B cell are inversely proportional to the affinity of the antigen it binds. Applying these two features, the AIS developed called CLONALG, which has been used to perform the tasks of pattern matching and multi-modal function optimisation. For example, in pattern matching, a set of patterns S, to be matched considered to be antigens. The task of CLONALG is to produce a set of memory antibodies, M, that match the members in S.

Clonal selection algorithms share many similarities with evolutionary algorithms [22], although importantly the selection and mutation mechanisms are influenced by the affinities of antibody-antigen matching [23]. Due to this similarity, many of the theoretical approaches applied to evolutionary algorithms are applicable to clonal selection algorithms also. Timmis et al., [24] summarizes much of the theoretical work done on AIS to date. This includes the work of Edward Clark et al. [25] who develop an exact Markov chain model of the clonal selection algorithm called the B-cell

algorithm (BCA) [26], proving its convergence. They go on to show how the model can be applied to give insight into optimal parameter settings for the BCA in a function optimization landscape. Other AIS that have been inspired by the adaptive immune mechanisms of B cells are AIRS [27], a supervised learning algorithm, and IA that has been used in numerous applications is studied [28].

Input: S = Set of patterns to be recognised,

n = the number of worst elements to select for removal

Output: M = set of memory detectors capable of classifying unseen patterns

begin.

Create an initial random set of antibodies, A

For all patterns in S do

Determine the affinity with each antibody in A
Generate clones of a subset of the antibodies in A with the highest affinity.

The number of clones for an antibody is proportional to its affinity

Mutate attributes of these clones inversely proportional to its affinity.

Add these clones to the set A, and place a copy of the highest affinity antibodies in A into the memory set, M

Replace the n lowest affinity antibodies in A with new randomly generated antibodies

End

End

Figure I: CLONALG algorithm for pattern recognition

III. EXPERIMENTAL SETUP AND RESULTS

The KDDCUP'99 data set was created by processing the tcp dump portions of the 1998 DARPR Intrusion Detection System (IDS) evaluation dataset, created by Lincoln Labs, U.S.A. They acquired nine weeks of raw tcp dump data. This was processed into about five million connection records. The data set contains a total of 24 attack types (connections) that fall into 4 major categories: Denial of service (Dos), Probe, User to Root (U2R), Remote to User (R2L). Each record is labelled either as normal, or as an attack, with exactly one specific attack type. In this paper we use the corrected KDD dataset available for research and used 20% of the data with two class labels normal and anomaly. Table I and II lists the detailed Accuracy by class of CLONALG and IMMUNOSI

Table I : Detailed Accuracy by Class of CLONALG

TP Rate	FP Rate	Precision	Recall	F measure	Class
0.383	0.12	0.407	0.383	0.395	normal
0.876	0.61	0.865	0.876	0.871	anomaly

Table II : Detailed Accuracy By Class: IMMUNOS 1

TP Rate	FP Rate	Precision	Recall	F measure	Class
0.86	0.23	0.444	0.86	0.586	normal
0.76	0.14	0.961	0.761	0.85	anomaly

Table III : Detailed Accuracy By Class: Naïve bayes

TP Rate	FP Rate	Precision	Recall	F measure	Class
0.83	0.38	0.326	0.83	0.468	normal
0.61	0.17	0.943	0.618	0.747	anomaly

Table IV : Evaluation on test split Summary

SUMMARY	**CLONALG**	**NAIVE BAYES**	**IMMUNO**
Correctly Classified Instances	78.6667%	65.6793%	77.9325%
Incorrectly Classified Instances	21.3333%	34.3207%	22.0675%
Kappa Statistic	0.2653	0.2798	0.4555
Mean Absolute Error	0.2133	0.3343	0.2207
Root Mean Squared Error	0.4619	0.562	0.4698
Relative Absolute Error	71.7605%	112.4401%	74.2301%
Root Relative Squared Error	119.8079%	145.7894%	121.852%
Total Number of Instances	11850	11850	11850

IV. CONCLUSIONS

In this paper we used a subset of the NSL - KDD dataset to avoid the problem of the classification algorithms leaning more towards the frequent item set. The dataset used contained more of anomalous records compared to normal records. Clonalg was able to detect the anomalous packets better than either Naive bayes or IMMUNOS. Only 12% of anomalous packets were predicted as normal packets compared to Naive Bayes which predicted 38% of anamolous packets as normal packets. For network intrusion detection system the choice between false rejection rate and false acceptance rate is to be decided by the system administrator based on the security requirement of the network.

REFERENCES

[1] Hui Wang, Guoping Zhang, Huiguo chen and Xueshu Jiang, "Mining Association Rules for Intrusion Detection",2009 IEEE International conference on frontier of Computer Science and Technology.
[2] S. Northcutt and J. Novak, "Network Intrusion Detection:An Analyst's Handbook," 2nd Edition, New Riders Publishing,Berkeley, 2000.

[3] Christoph Ehret, 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.

[4] Zhu, Dan, Data mining for network intrusion detection: A comparison of alternative methods Publication: Decision Sciences, 2001.

[5] George S.Oreku, Fredrick J. Mienzi, " Intrusion Detection Based on Data Mining,2009 ,Eighth IEEE International Conference on Dependable, Autonomic and Secure computing.

[6] http://www.spamlaws.com/how-data-mining-helps-intrusion-detection.html

[7] Steven Molnar , M. C., David Ellsworth , Henry Fuchs (1994). A Sorting Classification of Parallel Rendering, IEEE Computer Graphics and Applications. *IEEE Computer Graphics and Applications, 14*(4), 23-32.

[8] Eric Bloedorn, Alan D. Christiansen, William Hill, Clement Skorupka, Lisa M. Talbot, Jonathan Tivel, Data Mining for Network Intrusion Detection: How to Get Started, This work was sponsored by the MITRE Technology Program as a MITRE Sponsored Research (MSR) Project.

[9] Sophia Kaplantzis, Nallasamy Mani, a study on classification techniques for network intrusion detection

[10] Dewan Md. Farid, Jerome Darmont, Nouria Harbi, Nguyen Huu Hoa, and Mohammad Zahidur Rahman," Adaptive Network Intrusion Detection Learning: Attribute Selection and Classification, World Academy of Science, Engineering and Technology 2009

[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 junzhong huang houkuan , An evolving intrusion detection system based on natural immune system proceedings of IEEE TENCON'02

[14] John M. Hall ,an investigation into immune-based intrusion detection, December 2003, University of Idaho.

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

[16] 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.

[17] 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.

[18] D.Dasgupta, Z.Ji F.GonZalaz, IEEE, 2003, The University of Memphis.

[19] de Castro, L. N.; Von Zuben, F. J. (2002). "Learning and Optimization Using the Clonal Selection Principle" (PDF). *IEEE Transactions on Evolutionary Computation, Special Issue on Artificial Immune Systems* ,IEEE (3):239–251.

[20] Khaled A, Al.Sheshtawi,H.M.Abdul Kader, Nabil A.Ismail, Artificial Immune Clonal Selection Algorithms: A Comparative Study of CLONALG, opt-IA, and BCA with Numerical Optimization Problems, IJCSNS International Journal of Computer Science and Network Security, VOL.10 No.4, April 2010.

[21] Forrest, S.; Perelson, A.S.; Allen, L.; Cherukuri, R. (1994). "Self-nonself discrimination in a computer" (PDF). *Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy*. Los Alamitos, CA. pp. 202–212.

[22] Mingxi Wu, Christopher Jermaine, 'A Bayesian Method for Guessing the Extreme Values in a Data Set', Department of Computer and Information Sciences and Engineering, University of Florida, Gainesville, FL, USA.

[23] 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.

[24] Timmis, J. and M. Neal (2001), 'A Resource Limited Artificial Immune System'. Knowledge Based Systems 14(3/4), 121–130.

[25] Edward Clark, Andrew Hone, and Jon Timmis, "A markov chain model of the B-cell algorithm," Artificial Immune Systems, pp. 318-330, 2005.

[26] Tom M. Mitchell (2005), 'Generative and discriminative classifiers: naive bayes and logistic regression', Lecture notes on Machine learning.

[27] Watkins, A. (2001), 'A Resource Limited Artificial Immune Classifier'. Master's thesis, Mississippi Sate University.

[28] Mingxi Wu, Christopher Jermaine, 'A Bayesian Method for Guessing the Extreme Values in a Data Set', Department of Computer and Information Sciences and Engineering, University of Florida, Gainesville, FL, USA.