Volume 2, No. 6, Nov-Dec 2011

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

AIS : A Computational Approach for Network Intrusion Detection System

R.Sridevi*
Dept. Of Information Technology
Shri Angalamman College of Engg & Tech
Tiruchirappalli, India
devivelon@yahoo.com

Dr.Rajan Chattemvelli
Dept. Of Information Technology
Periyar Maniammai University
Tanjore, India

Abstract: Computer systems today can be seen as a problem of pattern classification, the system must deal with some intrinsic characteristics that make it very difficult to detect intrusions directly using classical pattern recognition methods. For example, normal and anomalous states are distinguished using features that are multi-dimensional, and there is extreme asymmetry and high overlap in the amount of data available for these two systems of states. Furthermore, the patterns involved cannot be recognized by linear methods without kernel maps. The natural human immune system faces the same difficulties, but successfully protects the body against a vast variety of foreign pathogens. It is a self-adaptive and self-learning classifier that can recognize and classify threats by learning, long-term memory, and association. We have adapted the mechanisms of the human immune system to construct an intrusion detection system to protect computer networks.

Keywords: Artificial immune system, Network intrusion detection, Classification, Immunology.

I. INTRODUCTION

Rapid decision–making involves the processing of information in an effective way and this comprises of fast access, analysis, and interpretation of enormous amounts of data. The information managers are becoming increasingly reliant on the speedy retrieval and processing of multidimensional or spatial data. To perform rapid tabulation, visualization of data from different sources are now possible by means of faster processors and correlation manipulation of data from different sources has allowed additional information to be incidental that may be difficult to obtain directly. The inexpensive usages of computer networks have worsened the problem of illegal access and fraud of data. Increased connectivity not only provides access to larger and varied resources of data, but it also provides an access path to sensitive data from virtually anywhere on the network [1].

Internet security is a trendy and vigorous field; the attack patterns can changes from one year to the next in a very fast manner. There are several innovative ideas, such as "networks can be secured by means of encryption" and "networks can be secured by firewalls" or a combination of both. The most excellent place to start the inflections of these notions may be to look at the most common attacks. Obviously, many attacks are available in the real world as network hacking, when they are in fact done in more conventional ways [2].

As Internet candidness increase very quickly, more and more organizations are becoming at risk to a wide variety of cyber threat attacks. According to a recent survey by CERT/CC (Computer Emergency Response Team/ Coordination Center), the rate of attacks has been doubling every year in recent times. It has become very essential to make our information systems, especially those used for risky works in the military and commercial sectors, resistant to and tolerant of such attacks. By keeping up-to-date information with current affairs in computing can indicate numerous cases of attacks made on computer servers of well-known companies. The computer attacks may range from denial-of-service attacks to extracting personal or financial data like credit-card details [3].

Organizations are increasingly dependent on online information that was considered to be one of the important assets. Since more information is being stored and processed on network-based systems, it is highly required to ensure the confidentiality. The wide use of e-commerce has increased the need of protecting the system to a very high end. Network Intrusion Detection has become an integral part of the information security process [4].

Kazienko et al. describe several activities that cause threat to the safety of computer resources from unauthorized access to a specific computer or address domain [5]. Intrusions were broadly classified into internal attacks, external attacks and remote attacks.

Mamta et al. analyzes the network data and their signatures and found that the potential possibility of an unauthorized [6] usage attempts to:

a. Access private information
b. Manipulate information
c. Render a system unreliable or unusable

Being intrusive is when someone is forcing themselves into a situation or a place where they are not openly welcome. This can also be true when anyone is invading someone's own space by asking personal or direct questions.

Intrusion detection is the process of supervising the events occurring in a computer system or network and analyzing them for signs of possible attacks, which can violate computer security policies, acceptable use policies, or standard security practices. Hackers try to gain access to systems from the Internet, whereas authorized users of systems try to misuse their privileges or attempt to gain additional privileges. Although many incidents are malicious in nature, many others

are not; for example, a person might mistype the address of a computer and accidentally attempt to connect to a different system with or without authorization. Starbranch et al. describes various malicious activies including hacking, attempts by unauthorized users to damage your computer by examining network traffic, denial of service attacks or port scans [7].

Scarfone et al. has extensively dealt with challenges faced by the network against intrusion. He defines IDS as software that analyzes and automates the intrusion detection process in an effective way [8]. The IDS checks all possibilities in and around the network activity, and identifies suspicious signatures that may indicate a network or system attack that compromise a system. Keeping up-to-date information on current affairs in computing can confirm many types of attacks made on computer servers of well-known companies. A firewall is a useful, often essential defense, but current firewall technology is insufficient to detect and block all kinds of attacks [9]. Some of the functions of Intrusion detection systems are

a. Monitoring and analyzing user and system activity,
b. Assessing the integrity of critical systems and data files,
c. Recognizing activity patterns reflecting known attacks,
d. Responding automatically to detected activity, and
e. Reporting the outcomes of the detection process.

Data mining algorithms have been proposed for intrusion detection systems due to their adaptive and self learning capabilities. Zhu et al tested several data set and detect the success rate of network intrusion detection systems based on size of feature of data that increases the prediction and accuracy both false alarm rates and detection efficiency [10].The security of our computer systems and data are at continuous risk. The extensive growth of Internet and availability of hacking tools has increased attacks on networks even by novice users. This has imposed an additional critical component in monitoring for the network administrator. The highlighted points below are the areas in which data mining technology may be applied or further developed for intrusion detection:

a. Data mining classification algorithms for intrusion detection.
b. Association and correlation analysis, and aggregation to help select and build discriminating attributes
c. Analysis of stream data
d. Distributed data mining
e. Visualization and querying tools

Data Mining System Products and Research Prototypes [11] should be assessed based on the following multiple features:

a. Data types
b. System issues
c. Data sources
d. Data mining functions and methodologies.
e. Scalability
f. Visualization tools
g. Data mining query language and GUI.

Challenges faced in data mining algorithms are the requirement of large training sets to evaluate the algorithm.

Typically upward of 60% is used as training data and the remaining as test data. In this paper we investigate the Artificial immune system as classifiers and their effectiveness in using a small training set. This paper is organized as follows. Section II gives a brief introduction to Artificial Immune System (AIS); Section III describes the dataset and the experimental setup with the obtained results.

II. ARTIFICIAL IMMUNE SYSTEM

A novel technique called Artificial Immune Systems (AIS), inspired by immunology, has been proposed recently. These are systems and algorithms that can use the human immune system (HIS) as an inspiration. The HIS is a robust, decentralized, error tolerant and adaptive system in nature. Immune system is an inherent part of each and every living organism and has to adapt continuously for the survival of the species. Such properties are highly desirable for the development of novel computer systems.

Greensmith et al analyze the Artificial Immune System (AIS) in contrast to the available bio-inspired techniques, such as genetic algorithms and neural networks and found that the field of AIS covers a spectrum of algorithms that exist because different algorithms implement different properties of human cells [12]. The concepts of Artificial Immune System have been extracted and applied for solution to real world problems [13].

Dasgupta et al describes the similarities and differences between the Artificial Neural Network (ANN) system and Artificial Immune System (AIS). AIS is now gaining its popularity with several concepts from immunology extracted and applied for the solution of real-world science and engineering problems like theoretical modeling and simulation with a wide variety of applications [14]. The Human Immune System (HIS) is a rich source of theories. It can act as an inspiration for the creation of novel approaches to computational problems; this field of research is referred to as Immunological Computation (IC) [15].

The Artificial Immune System is highly distributed, highly adaptive, self-organizing in nature, maintains a memory of past events, and has the ability to learn about new encounters. It consists of various intelligent methodologies that can be used to churn out effective solutions to real world problems [16]. Timmis et al. have successfully implemented classification algorithms using Artificial Immune System with fairly good results. Immunos-81 a classification system was tested using two important data sets which are obtained from the medical field. Immunos-81 achieved an average classification rate of 83.2% on the Cleveland data set and approximately 73.5% on a second data set, which uses the first data set as the training data set. [17].

The AIS protects a complex system against malicious defects, thereby achieving its survival policy by an extension of the concept of organization of multi-cellular organisms to the information systems. The main features of AIS are self-maintenance, distributed and adaptive nature of the computational systems [18]. Although AIS is used for pure optimization, it can be made to work; this is probably the

missing point. AIS are powerful when a population of solutions is essential either during the search, or as an outcome. The concept of matching is involved in finding the solution. Since AIS are evolutionary algorithms, they are more suitable for problems that change over time, and need to be solved again and again, rather than one-off optimizations [19]. Table I shows the differences between HIS and AIS.

A supervised learning system, AIRS concentrates on the class of each antigen (feature vector in the training set) while generating ARBs that respond to the antigens. The most reactive ARBs generated are promoted to the memory cell pool which is the ultimate classification tool that remains after training is complete. The mutation built into the algorithm results in ARBs that are winning in part by not being identical to training vectors, while being analogous enough to later training vectors to be highly competitive. This is the source of AIRS's ability to simplify from the data. AIRS "grows" the memory cell pool from an initial size set by the user and seeded by training vectors. Typically, the number of cells that AIRS creates in the memory cell pool is about half the number of training cells presented to it. Once training is complete, the memory cells are used as a k-nearest neighbor classification system for test data [20].

Table I : Difference between HIS and AIS

	Human Immune System	**Artificial Immune System**
Integrity	This is a way to guarantee that the genetic codes present in the cells will not be corrupted by any pathogen.	Data has to be protected from intentional or accidental corruption.
Availability	This aspect allows the body to continue working even under attack of the pathogen.	Information such as the computer must be accessible when necessary and as desired.
Correction	The Mechanism prevents the immune system against attack of the (body) cells.	False alarms from an incorrect classification of computational events must be minimized.
Accountability	These are means adopted by the immune system identify, find and destroy the pathological agents.	The security system must be configured to preserve sufficient information from the intrusion that can be permitted to trace the origin of the attack.
Confidentiality	There are no concepts of secret data or no concepts of secret data or confidential information.	Data access must only be allowed to authorized users.

III. DATASET USED, EXPERIMENTAL SETUP AND RESULT

To evaluate our proposed method, we use the UDP data available in the KDD 99 dataset. The UDP stream contains three different types of intrusions namely teardrop, satan and nmap. As the goal of this work is to evaluate the efficiency of the AIRS algorithm using small training sets, we split the dataset into 15:85, 25:75, 30:70 ratios for training and test set data. Table II summarizes the training summary used in this work.

Table II : Training summary

Affinity Threshold	0.227
Total memory cell replacements	36,033
Mean ARB clones per refinement iteration	51.379
Mean total resources per refinement iteration	126.365
Mean pool size per refinement iteration	69.682
Mean memory cell clones per antigen	19.603
Mean ARB refinement iterations per antigen	2
Mean ARB prunings per refinement iteration	53.68

The classification accuracy of AIRS algorithm for 15% , 25% and 30% training set is tabulated in table III.

Table III: Classification accuracy

	25% training set	15% training set	30% training set
Correctly Classified Instances	99.47 %	99.12%	99.66 %
Incorrectly Classified Instances	0.531 %	0.88 %	0.34 %

The detailed accuracy by class and confusion matrix is provided in Table IV. It is seen that recall starts decreasing as the training set decreases . As teardrop type of attacks are very sparsely represented in the total dataset, it contributes to the maximum errors.

Table IV: Detailed accuracy by class

25% training set				
TP Rate	FP Rate	Precision	Recall	Class
1	0.214	0.995	1	normal.
0.327	0	1	0.327	teardrop.
0.984	0	1	0.984	satan.
0.813	0	0.987	0.813	Nmap
15% training set				
0.994	0.14	0.997	0.994	Normal
0.639	0.005	0.399	0.639	Teardrop
0.983	0	0.994	0.983	Satan
0.829	0	0.972	0.829	nmap
30% training set				
1	0.132	0.997	1	Normal

0.584	0	0.985	0.584	Teardrop
0.988	0	0.996	0.988	Satan
0.884	0	0.956	0.884	nmap

IV. CONCLUSION

This paper investigated the AIRS classifier algorithm using smaller percentage of training data. The thumb rule for classification algorithm is to use at least 60% of the dataset as training dataset. This paper shows that AIRS performance does not get degraded even if a smaller percentage of training data is used to train the classifier algorithm. This can be very useful to detect newer types of attacks. As such, the AIRS represent a powerful example of learning within an adaptive, non-linear network, containing an explicit, content addressable memory, implemented in a relatively simple computer program. We also believe that the ideas encompassed by the immune system can provide a wealth of problem solving methods, which have yet to be fully realized.

V. REFERENCES

[1] Sandeep Kumar, Classification and detection of computer intrusions, PhD Thesis, Purdue University, 1995.

[2] Roger Needham and Butler Lampson, Network attack and defense security engineering: A Guide to building dependable distributed systems, 2000.

[3] Springer Verlag, Proceedings of the International Conference on Artificial Immune Systems (ICARIS), 2003.

[4] Norbik Bashah, Idris Bharanidharan Shanmugam, and Abdul Manan Ahmed, "Hybrid intelligent intrusion detection", World Academy of Science, Engineering and Technology, 11 2005.

[5] Przemyslaw Kazienko & Piotr Dorosz, Intrusion detection systems (IDS) Part I - (network intrusions; attack symptoms; IDS tasks; and IDS architecture), Apr 07, 2003.

[6] Mamata Desai, M.Tech Dissertation, Distributed intrusion detection, Department of Computer Science and Engineering, Indian Institute of Technology Bombay, 2003.

[7] Spencer Starbranch, Network intrusion detection systems to detect intruders. Article Source: http://EzineArticles.com/?expert=Spencer_Starbranch 2010.

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

[9] L de Castro, J Timmis, Artificial Immune Systems: A New Computational Intelligence Approach, 1st Edition., Springer Verlag, 2002.

[10] Zhu, Ming Ye , G. Premkumar, "An Evaluation of Size-based Traffic Feature for Intrusion Detection", Journal of Information System Security, JISSec 3 (1) 2007.

[11] http://www.dataminingtools.net/index.php dataminingtools.net

[12] Julie Green Smith, Amanda Whit Brook and Uwe Aickelin, Artificial Immune Systems, arXiv:1006.4949v1 [cs.AI,]2010.

[13] U. Aickelin, D. Dasgupta, F. Gu. Artificial Immune Systems , chapter 13. To appear in (updated) edited volume on Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques. 2010.

[14] Dipankar Dasgupta, Artificial Immune Systems: A Bibliography CS TECHNICAL REPORT No. CS-07-004 December 2007 Version 5.8.pg:126-137

[15] 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.

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

[17] Abbass, Hussein A. and Sarker, Ruhul A. and Newton, Charles S., eds Timmis, Jon and Knight, Thomas, (2001) *Artificial Immune Systems: Using the Immune System as Inspiration for Data Mining.* In:. Data Mining: A Heuristic Approach. Group Idea Publishing, Harrisburg PA, pp. 209-230. ISBN 978-1930708259.

[18] Artificial immune system-: An emergent technology for autonomous intelligent systems and data mining Springer Verlag, Proceedings of the International Conference on Artificial Immune Systems (ICARIS), 2003.

[19] Zhu, Dan, Data mining for network intrusion detection: A comparison of alternative methods Decision Sciences Volume 32, Issue 4, pages 635–660, December 2001 .

[20] de Castro, L. N. and Von Zuben, F. J., "Learning and Optimization Using the Clonal SelectionPrinciple". IEEE Transactions on Evolutionary Computation, Special issue on Artificial Immune Systems. 2002