A Result Evolution of An Artificial Immune System for Intrusion Detection System to Improve the Detection Rate

Pallvi Dehariya, Shiv K Sahu, Amit Mishra

Abstract: ***This paper presents an intrusion detection system architecture based on the artificial immune system concept. In this architecture, an innate immune mechanism through unsupervised machine learning methods is proposed to primarily categorize network traffic to "self" and "non-self" as normal and suspicious profiles respectively. Unsupervised machine learning techniques formulate the invisible structure of unlabeled data without any prior knowledge. The novelty of this work is utilization of these methods in order to provide online and real-time training for the adaptive immune system within the artificial immune system. The proposed intrusion detection system will use the concepts of the artificial immune systems (AIS) which is a promising biologically inspired computing model. AIS concepts that can be applied to improve the effectiveness of IDS.***

Keywords: ***Intrusion detection system, Artificial Immune system, clustering***

I. INTRODUCTION

Computer Security is used frequently, but the content of a computer is vulnerable to few risks unless the computer is connected to other computers on a network. As the use of computer networks, especially the Internet, has become pervasive, the concept of Computer security has expanded to denote issues pertaining to the networked use of computers and their resources. The major technical areas of computer security are usually represented by the initials confidentiality, integrity, and authentication or availability. "denial of service" attacks, which are sometimes the topic of national news, are attacks against availability. Other important concerns of computer security professionals are access control and no repudiation. The main goal of intrusion detection is to detect unauthorized use, misuse and abuse of computer systems by both system insiders and external intruders. Among automated intrusion detection systems, a particular system for network intrusion detection, known as a network-based intrusion detection system (IDS), monitors any number of hosts on a network by scrutinizing the audit trails of multiple hosts and network traffic. It is usually comprised of two main components: an anomaly detector and a misuse detector [1][2]. The anomaly detector establishes the profiles of normal activities of users, systems, system resources, network traffic and/or services and detects intrusions by identifying significant deviations from the normal behavior patterns observed from profiles. The misuse detector defines suspicious misuse signatures based on known system vulnerabilities and a security policy. This component probes whether these misuse signatures are present or not in the auditing trails. This paper proposes the use of negative selection and niching of artificial immune system for developing effective network-based IDS. An overall artificial immune model for network intrusion detection presented in (Kim and Bentley, 1999b) consists of three different evolutionary stages: negative selection, clonal selection, and gene library evolution. Among these stages, the first stage, negative selection, is investigated in this paper. We present a more efficient implementation of negative selection using a niching feature of artificial immune systems [9]

II. LITERATURE SURVEY

A lot of research works have been carried out in the literature for intrusion detection and some of them have motivated us to take up this research. Brief reviews of some of those recent significant researches are presented below:

Tich Phu oc Tran have applied Machine Learning techniques to solve Intrusion Detection problems within computer networks. Due to complex and dynamic nature of computer networks and hacking techniques, identifying malicious activities remains a challenging task for security experts, that is, defense systems that were currently available suffer from low detection capability and high number of false alarms.

Ye Yuan et proposed a method of evidence assignment in combination with Dempster-Shafer theory to identify network attack data. In this method, extracted features were identified by a multigeneralized regression neural network classifier, which determined the basic probability assignment.

Snehal A proposed the decision tree based algorithm to build multiclass intrusion detection system. Support Vector Machines was the classifiers which were initially designed for binary classification.

Shun J and **Malki H. A.** presented a neural network-based intrusion detection method for the internet-based attacks on a computer network.

Aida O. Ali id a relative study between the performances of recent nine artificial neural networks (ANNs) based classifiers was assessed centered on a particular set of features. The outcomes showed that; the Multilayer perceptrons (MLPS) based classifier yielded the best results; about 99.63% true positive attacks were detected.

Pohsiang Tsai suggested a Machine Learning (ML) framework in which various types of intrusions would be detected with different classifiers, containing different attribute selections and learning algorithms. Appropriate voting techniques were used to combine the outputs of these classifiers.

Revised Version Manuscript Received on June 13, 2016.

Pallvi Dehariya, M.Tech. Scholar, Department of Information Technology, Technocrats Institute of Technology, Bhopal (M.P.), India.

Dr. Shiv K Sahu, Professor & Head, Department of Information Technology, Technocrats Institute of Technology, Bhopal (M.P.), India.

Amit Mishra, Asst. Professor, Department of Information Technology, Technocrats Institute of Technology, Bhopal (M.P.), India.

The pattern-learning abilities of the IS has been modeled and described by **Timmis, Neal, and Hunt (2008) and Dasgupta, Cao, and Yang (2003)** who successfully applied their AISs to recognition and classification tasks.

Also **Byoung-Doo** in 2006 built IDS deals well various mutated attacks, as well as well-known attacks by using Time Delay Neural Network classifier that discriminates between normal and abnormal packet flows. It seems that the area where the notion of AIS has been most widespread is in the area of computer security.

A. H. M. Rezaul Karim proposed collaborative IDS for MANET using Bayesian method using a set of very useful features which guarantee the effectiveness of the IDS [12].

L. Khan and et al. proposed a method with a scalable solution for detecting network based anomalies [13]. They used Support Vector Machines (SVM) for classification. They used the Dynamically Growing Self-Organizing Tree (DGSOT) algorithm for clustering.

Tsong and introduced a three-tier architecture of intrusion detection system which consists of a blacklist, a whitelist and a multiclass support vector machine classifier.They designed a three-tier IDS based on the KDD'99 benchmark dataset.

Weiming Hu proposed an intrusion detection algorithm based on the AdaBoost algorithm. The discrete AdaBoost algorithm was selected to learn the classifier.

Hu Zhengbing1 proposed an algorithm to use the known signature to find the signature of the related attack quickly. They used nine different-sized databases,

Amit Kumar Choudhary proposed a neural network approach to improve the alert throughput of a network and making it attack prohibitive using IDS. For evolving and testing intrusion the KDD CUP 99 dataset were used.

Stefano Zanero proposed a novel architecture which implements a network-based anomaly detection system using unsupervised learning algorithms. They described how the pattern recognition features of a Self Organizing Map algorithm can be used for Intrusion Detection purposes on the payload of TCP network World Journal of Science and Technology 2012, 2(3):127-133 131 packets.

III. PROBLEM IDENTIFICATION

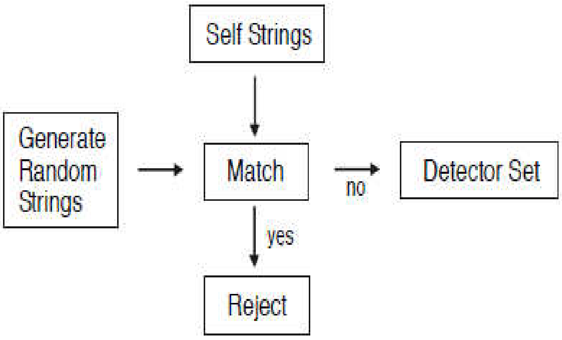
The main drawback of traditional methods is that they cannot detect unknown intrusion. Even if a new pattern of the attacks were discovered, this new pattern would have to be manually updated into system. It is also capable of identifying new attacks to some degree of resemblance to the learned ones, the neural networks are widely considered as an efficient approach to adaptively classify patterns [Boger][11], but their high computation intensity and the long training cycles greatly hinder their applications, especially for the intrusion detection problem, where the amount of related data is very important.

IV. PROPOSED APPROACH

The first negative selection algorithm was proposed by Forrest **et al** (1994) to detect data manipulation caused by a virus in a computer system. The starting point of this algorithm is to produce a set of self strings, S, that define the normal state of the system. The task then is to generate a set of detectors, D, that only bind/recognize the complement of S. These detectors can then be applied to new data in order to classify them as being self or non-self. The algorithm of Forrest **et al** produces the set of detectors via the process outlined in below.

4.1 Algorithm Overview

This work uses a negative selection algorithm to build an anomaly detector. This is achieved by generating detectors containing non-self patterns. The overview of this algorithm is provided in figure 4.1 and 4.2. The negative selection algorithm for network intrusion detection used 'Self' is built by profiling the activities of each single network connection**.**



The Procedure of Negative Selection algorithm is as follow
input S_{seen} = set of seen known self elements
output: D = set of generated detectors

```
Begin
Repeat
  • Randomly generate potential detectors and place them in a set $P$
  • Determine the affinity of each member of $P$ with each member of the self set $S_{seen}$
  • If at least one element in $S$ recognizes a detector in $P$ according to a recognition threshold,
        then the detector is rejected, otherwise it is added to the set of available detectors $D$
  • until Stopping criteria has been met
End
```

V. RESULT

In this experiment, we investigate computation time of Negative Selection Algorithm and K-Mean algorithm. In the training phase, the Selection Algorithm was used to cluster the training data. After training, each cluster was labeled according to the majority type of data in this cluster. For instance, if more than 50% of the connections in cluster were intrusions, the cluster and its centroid weight vector would be labeled as intrusion. Negative Selection Algorithm perform significantly better (p < 5%) than the others in terms of computation time with much less run time Comparing the results for 100 clusters is shown in table (6.6.1). Negative Selection Algorithm algorithms perform significantly better (p < 5%) than the others in terms of computation time. Comparing the results for 100 clusters, we observe that the K-Means take more execution time than Selection Algorithm.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

Table 1 : Clustering results with 100 clusters with time efficiency

Cluster	Algorithm	
	K-Mean	Negative Selection Algorithm
	Time (ms)	Time (ms)
20	70	65
40	89	75
60	95	85
80	102	95
100	108	101

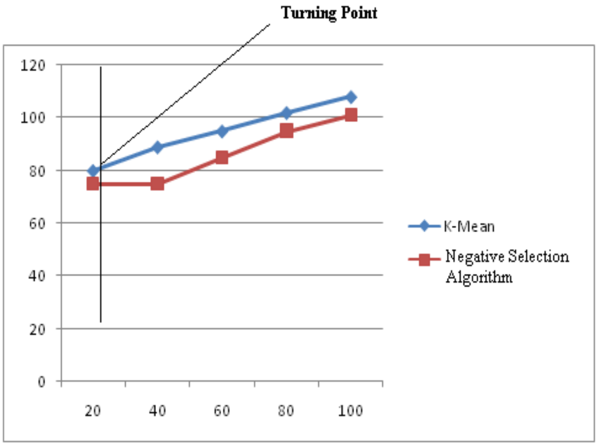


Fig: 1. The number of clusters vs. Computation Time

VI. CONCLUSION

This paper has described the Promising clustering and detection results encourage us to proceed our future work in several directions. Identifying the precise attack category associated with a cluster and the discriminating features that are unique to a given cluster can do a further detailed analysis of individual clusters.

In addition, feature selection/weighting for clustering will be investigated. This will eventually enhance our understanding and detection of new attack categories. Sophisticated self-labeling techniques, taking into consideration of additional network security domain knowledge, can be developed to improve the performance of clustering-based intrusion detection

REFERENCES

1. Cho, Sung-Bae. 2003. .Artificial Life Technology for Adaptive Information Processing. Chapter 2 in Future Directions for Intelligent Systems and Information Sciences: The Future of Speech and Image Technologies, Brain Computers, WWW, and Bioinformatics, edited by Nikola Kasabov, Volume 45 of Studies in Fuzziness and Soft Computing, 13.33. Heidelberg, Germany: Physica-Verlag. ISBN 3-7908-1276-5.
2. Dasgupta, Dipankar. 1999, October. .Immunity-Based Intrusion Detection System: A General Framework.. Proceedings of the 22nd National Information Systems Security Conference (NISSC). National Institute of Standards and Technology and National Computer Security Center, Hyatt Regency.Crystal City, Virginia, United States.
3. Dasgupta, Dipankar, Yuehua Cao, and Congjun Yang. 2003, July 13.17. .An Immunogenetic Approach to Spectra Recognition.. Edited by Wolfgang Banzhaf, Jason Daida, Agoston E. Eiben, Max H. Garzon, Vasant Honavar, Mark Jakiela, and Robert E. Smith, Proceedings of the Genetic and Evolutionary Computation (GECCO) Conference, Volume 1. Orlando, Florida, United States: Morgan Kaufmann, 149.155. ISBN 1-55860-611-4.
4. Dasgupta, Dipankar, and Stephanie Forrest. 1996, June 19.21. .Novelty Detection in TimeSeries Data using Ideas from Immunology.. Proceedings of the 5th International Conference on Intelligent Systems. Reno, Nevada, United States.
5. Nong Ye and Xiangyang Li. A scalable clustering technique for intrusion signature recognition. In Proc. 2nd IEEE SMC Information Assurance Workshop, pages 1-4, 2001.
6. Yu Guan, Ali A. Ghorbani, and Nabil Belacel. Y-means: a clustering method for intrusion detection. In Canadian Conference on Electrical and Computer Engineering, pages 1-4, Montral, Qubec, Canada, May 2003.
7. Teuvo Kohonen. Self-Organizing Map. Springer-Verlag, New York, 1997
8. J. D. Banfield and A. E. Raftery. Model-based Gaussian and non-Gaussian clustering.
9. FAQ: Network Intrusion Detection Systems, Version 0.8.3, March 21, 2000 [Intrusion Detection
10. I.T. Jolliffe. Principal Component Analysis. Springer-Verlag, New York, 1989.
11. Kohonen, T. 1995. Self-Organizing Maps, volume 30 of Springer Series in Information Sciences. Berlin, Heidelberg: Springer. (Second Extended Edition 1997).
12. Leonid Portnoy, “Intrusion Detection with Unlabeled Data using Clustering", Undergraduate Thesis, Columbia University, New York, NY, Dec. 2000.
13. Lane, T., and Brodley, C. E. 1999. Temporal sequence learning and data reduction for anomaly detection. ACM Transactions on Information and System Security 2(3): 295—331.
14. Michael Sobirey's Intrusion Detection Systems http://www.rnks.informatik.tucot.
15. “NIST Special Publication on Intrusion Detection Systems“, SP 800-31 Computer Security Resource Center (CSRC), National Institute of Standards and Technology (NIST), Nov. 2001, p.15.
16. P.Lichodzijewski, A. n. Zincir-Heywood and M. I. Heywood, “Host-based intrusion detection using Neural Gas,” Proceedings of the 2002 IEEE World Congress on Computational Intelligence, 2002 (in press).
17. Salvatore J. Stolfo, Wei Fan, Wenke Lee, "Cost-based Modeling for Fraud and Intrusion Detection: Results from the JAM Project", Proceedings of the 2000 DARPA Information Survivability Conference and Exposition, 2000.
18. Vesanto J., Alhoniemi E., “Clustering of the Neural Gas Map,” IEEE Transactions on Neural Networks, 11(3), pp 586-600, 2000
19. Wenke Lee, Sal Stolfo, and Kui Mok. Mining in a data environment: Experience in network intrusion detection. In Proc. 5thACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, pages 114{124, San Diego, CA, August 1999.
20. Wenke Lee and Sal Stolfo, “Data Mining Approaches for Intrusion Detection”, Proceedings of the Seventh USENIX Security Symposium (SECURITY '98), San Antonio, TX, January 1998.
21. Wei Fan, Wenke Lee, Sal Stolfo, and Matt Miller (2000) ``A Multiple Model CostSensitive Approach for Intrusion Detection", Eleventh European Conference on Machine Learning (ECML '00) 2000.
22. Wei Fan, Matt Miller, Sal Stolfo, Wenke Lee, and Phil Chan, “Using Artificial Anomalies to Detect Unknown and Known Network Intrusions”, CA, November 2001

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.