International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 12– No.30, June 2020 – www.ijais.org

A Novel Immune Inspaired Concept with Neural Network for Intrusion Detection in Cybersecurity

Adeniji Oluwashola David
Department of Computer Science
University of Ibadan

Ukame James Joseph
Department of Computer Science
University of Ibadan

ABSTRACT
Artificial immune system (AIS) that depicts the way the human immune system (HIS) responds to threats or attacks in the body . AIS was used by researchers to solve intrusion problems.Immune system algorithms like the clonal selection theory, immune networks, negative selection algorithms and danger theory concepts, although has achieved some level of results, but not adequate especially in the cybersecurity domain. In this study a model based on AIS concepts that will find a significant application in cybersecurity was developed.The negative selection algorithm (NSA) which is a class of very flexible algorithm will divide the problem space into self and non-self which was used to build the model. The detector generation phase of the NSA was improved and a neural network technique was incorporated to build the model. The developed model called NNET NSA (Neural Network Negative Selection Algorithm) used the NSLKDDCup1999 dataset to test the model. An R script was written using the R programming language and implementation was done on both Rstudio and Rapid Miner environments.Experimental results showed that the model NNET NSA achieved a high classification accuracy of 90.1% within a computation time of 15seconds as compared with two classification algorithms; support vector machine (SVM) and Naïve Bayes which achieved a classification accuracy of 65.01% and 81.66% within a computation time of both 215.81seconds and 100.15seconds respectively on the R console. The developed model (NNET NSA) further showed a low wrong classification of 3.9% as compared with SVM; 4.8% and Naive Bayes; 4.2% respectively.

Keywords
Artificial Immune System, Artificial Neural Network, Cybersecurity, intrusion detection

1. INTRODUCTION
The immune network theory hypothesizes the activities of the immune cells, the emergence of memory and the discrimination between our own cells (known as self) and external invaders (known as non-self). It also suggests that the immune system has an internal image of all pathogens (infectious non-self) to which it was exposed during its lifetime.Idiotypic immune network theory, initially proposed by Jerne in 1974. The theory postulates that interactions between immune cells (and not necessarily external agents) cause modulation in the behaviour of the immune system as a models have been developed, although no physical evidence exists to support the theory [1].

2. RELATED WORKS
Immune networks based on AIS has also been seen to gain numerous application in clustering and filtering crude data sets described by high-dimensional samples [2].Artificial immune system mimics the way the human or biological immune system acts and responds to threats in the body. AIS has been applied in numerous field by several researchers like [3]developing an algorithm for solving a multi-criteria customer allocation problem in supply chain environment,[4]developing an IDS based on the concept of an AIS as shown in figure 1.

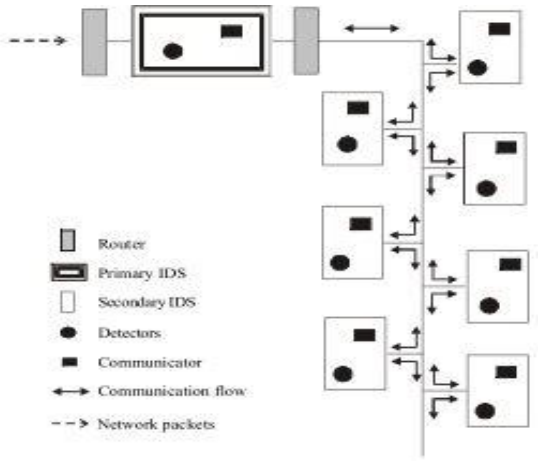

Fig 1: Architecture of an AIS Kim and Bentley, 1999b

Abstracting the functionality of the HIS [Human Immune System], AIS seek to explore such concepts and algorithms to solve computer security problems, network intrusion detection issues, etc. either in real time or experimentally simulated.

An Artificial Neural Network (ANN) is based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit a signal from one artificial neuron to another. An artificial neuron that receives a signal can process it and then signal additional artificial neurons connected to it.

2.1 Cybersecurity
Cybersecurity according to ISO [5] has been defined as the preservation of confidentiality, integrity and availability of information in the cyberspace", with an accompanying definition of cyberspace as "the complex environment resulting from the interaction of people, software and services on the Internet by means of technology devices and networks connected to it, which does not exist in any physical form.

Cyber security must address not only deliberate attacks, such as from disgruntled employees, industrial espionage, or terrorists, but inadvertent compromises of the information infrastructure due to user errors, equipment failures, and

natural disasters. (Yan, Qian, Sharif, & Tipper, 2012). Vulnerabilities might allow an attacker to penetrate a network, gain access to control software, and alter load conditions to destabilize normal business processes in unpredictable ways.The prediction of incoming attacks is achieved in a timely manner which enables security professionals to install defense systems in order to reduce the possibility of such attacks [7]in Zero Day attack Prediction.

Countermeasures which prohibit and tries to defend systems from cyberattack have been continuously developed. Some countermeasures are designed to limit physical access to an area which permits access to the IT infrastructure.The internet service driven network is a new approach to the provision of network computing that concentrates on the services you want to provide as adopted in [8].These include key entry systems, retinal Mechanisms.or fingerprint scans, and armed guards. Other countermeasures are designed to block access and/or protect privacy over networks serving the organization.

and spyware scanners.There are many efforts under way to stop the increase of spam that plague almost every user on the internet. [9]. Also, some countermeasures are designed to permit recovery if an intrusion is successful, such as backing up important files on a frequent basis.Access Attacks where an intruder gains access to a device to which he has no right to access. Denial of Service (DOS) which is an intrusion into a system by disabling the network with the intent to deny service to authorized users. DOS attacks make a computing or memory resource too busy or too full to handle legitimate requests, thus denying legitimate users access to a machine. The Multi-layered cyber defense with feed forward and feedback signaling mechanisms is shown in figure 2 below.

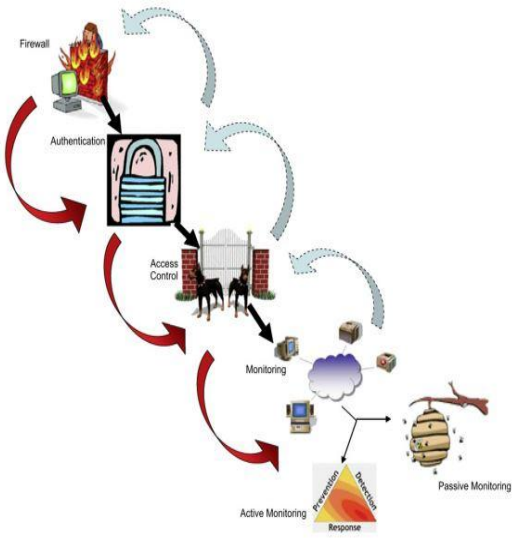

Fig 2: Multi-layered cyber defense with feed forward and feedback signaling mechanisms Dasgupta,2007

Different methods have been employed to secure and protect the shared and sensitive data. The significant roles of encryption algorithms are numerous and essential in information security [10] ***in*** Comparative Study of Symmetric Cryptography Mechanism .

3. METHODOLOGY

The Developed Novel Model was coded in the R programming language with the following modules:

Input module.(ii).Network Decoder module (optional module as seen from the methodology) (iii).Negative selection module.(iv) Classification module. The network traffic takes data from the input module to the negative selection module which then passes such traffic to the classification module for proper classification into either self or non-self-profile.

The input module captures the traffic from any given network interface from figure 3. The NetPcap library is used for this purpose. PCAP files are generated and stored in the local directory. Optionally, if the user does not want the traffic to be captured from the network interface, the user could directly feed in the PCAP files. Usually two types of files are to be fed into the input module namely the self-file and the test file. The self-file is the training file that will be used by the algorithm for training and generation of detectors. The test file is the packet capture of the normal traffic that is to be monitored. Traffic in this file will be classified as normal or anomalous in the end.

The negative selection module uses the self and test strings generated from the previous module to generate detectors which will then be used in classification. The working of negative selection module is explained in detail in the following section. The negative selection module generates set of detectors based on the self-strings. The detector set is represented using prefix directed acyclic graphs.

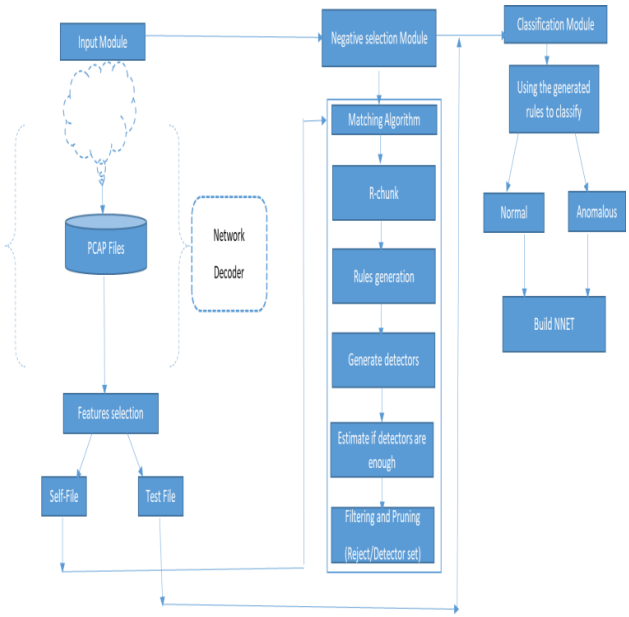

Fig 3.: The Novel Model of NNETNSA Original Feb., 2019.

These detectors are then used in classification module. The classification module is the final module that decides whether the traffic is self or non-self. The detectors are generated in such a way that it depicts traffic that is normal. So if the classification module finds a string in the test file that does not match with any of the detectors, it means that the string points to a traffic that is an anomaly. An anomaly score is generated depending on the level up to which the self and detector strings match, to facilitate in classification. If the anomaly score crosses a predefined and configurable anomaly threshold, then the case would be classified an anomaly. The anomaly threshold value can be set low if the network environment

operates under high risk levels. This low threshold value will ensure that even the slightest anomaly is flagged. Under normal network environments, it is not advisable to keep the anomaly threshold low, as it might lead to higher false positives and redundant alarms.The NSL-KDD dataset was used in this study. Such datasets served as a benchmark for the various parameters of an IDS like false positives, false negatives and detection rates.The Algorithm for the construction of prefix DAG is shown below.

Algorithm For: construction of prefix DAG

Detect_set_gen (S, x, r_s)

S: self-sample

X: number of detectors

r_s: self-radius

1. Let D be from an empty space
2. repeat
3. t be random samples from the space $[0,1]^n$
4. Repeat for all s_i C S …….{i=1,2,3…}
5. d← r-chunk
6. for every *l-r+1*
7. do
8. while d ≤ r_s, repeat step 2
9. D←D U t
10. until D=m
11. return D

Fig 4: Algorithm For NNET Detection Generation

The Human Immune System was mapped with the developed to Intrusion Detection Model.The mapping is shown below:

Body: The entire IDS Self-cells: Normal profiles

Non self-cells: Anomalous profiles

Antigen: sequences of observed protocol events recognized in packet headers or data path triples like is_guest_login, land, logged_in, src_bytes, num_failed_login, etc.

Antibody: A pattern with a predefined format.Chemical binding: binding of antibody to an antigen by use of a matching function or algorithm, r-chunk in this case.Note that an antibody will match an antigen if the antibody as a 1 in every position where the antigen has a 1 also. Or binding will occur at the position *i*, which a particular search bit occurs.

4. RESULT AND DISCUSSION

The implementation was performed on the RStudio development platform. Further experimental analysis were carried out on the model and evaluation of the developed model was made with two other classifier algorithms namely; Support Vector Machine (SVM) and Naïve Bayes.To load the NSLKDDCup dataset for artificial immune system for intrusion detection [12,13,14,15], we execute the command by typing the file name "NLC KDD+_20Percent.arff"" on the R console. The "NLC KDD+_20Percent.arff"" initiates the execution of the negative selection algorithm by loading the pre-processed dataset. This is depicted in figure 4.0 below showing the negative selection algorithm program loading intrusion detection dataset.

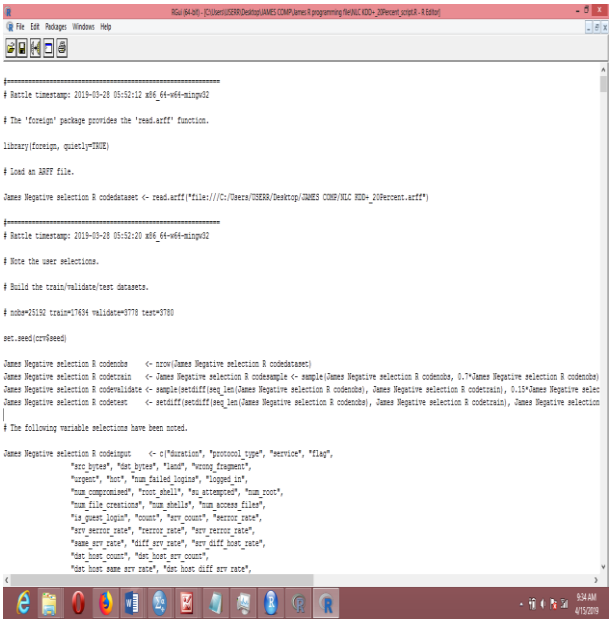

Figure 4.0:Loading NSLKDD dataset on the R console

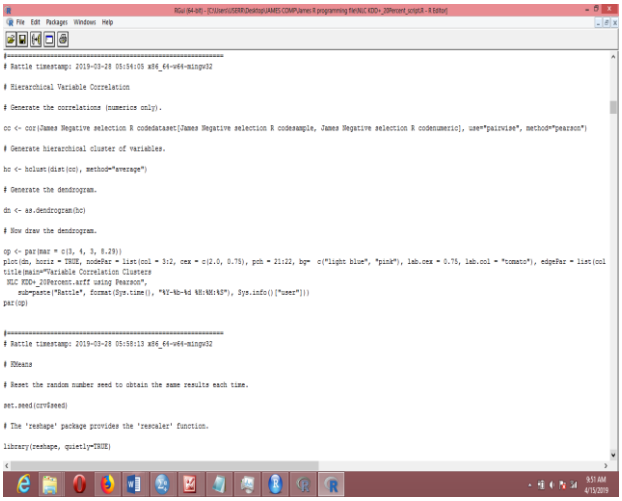

Figure 4.1: r-chunk matching of the antigens to create antibodies.

A one hundred and thirteen, ten (10) cluster sized, one (1) network (for optimal result) was built in the first instance with one thousand, two hundred and sixty four weights, each weight assignable to a node in the network. Performance (expected cost normalized to be between 0 and 1) is plotted on the y-axis. Operating points are plotted on the x-axis after being normalized to be between 0 and 1 by combining the parameters defining an operating point in the following way:

$$PCF(+) = \frac{p(+)c(-|+)}{P(+)c(-|+) + p(-)c(+|-)} \quad .(Equation1)$$

Where:

c(-|+) = cost of misclassifying a positive profile as negative

c(+|-) = cost of misclassifying a negative profile as positive

p(+) = probability of a positive profile and

p(-) = 1 – p(+).

Cost curves were defined to allow performance to be read off directly for any given operating point(Drummond & Holte, n.d.), for example, the performance when misclassification costs are equal and the two classes are equally likely can be read off the plot by looking at the cost curve's value at x = 0.5. A across a range of operating points, the performance which is between 0.28-0.65 (except when PCF (+) > 0.9) does not vary very much, hence, this is an optimal result.

There are 25,192 input nodes for self-data used in the algorithm. 17,634 (70% of the total input) data were used to train the model, 3,780 were used to test it, and 3,778 were used as validation. The accuracy and the computation time obtained were 90.1% and 15.02 seconds respectively. Fig 4.2 shows the cost curve.

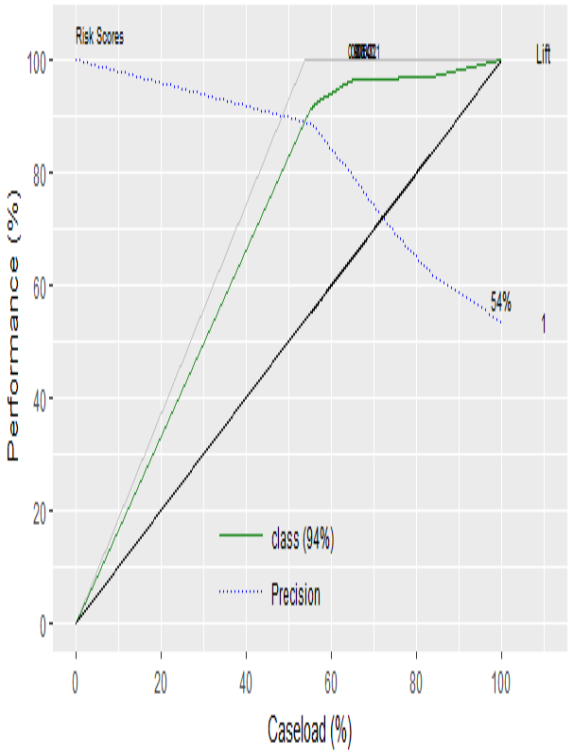

Fig 4.2: Risk curve obtained using NNET NSA.

Figure 4.2 above reveals that when combining the normal string and attack string, the percentage of correct classification of attack detected is 94%. That is NNET NSA correctly classified intrusion detected 94 times.

Figure 4.3 below shows the binary confusion matrix obtained by the system. The result portrays a total instance value of 25,192, a 13,449 of these instances were classified as normal profiles whereas 11,743 of such instances were detected as anomalous profiles.

```
=== Confusion Matrix ===

     a     b   <-- classified as
 13449     0 |     a = normal
 11743     0 |     b = anomaly
```

Figure 4.3: Confusion Matrix of NNET NSA

Table 4.1 below reveals that NNET NSA correctly classified 85.7% of the cyber-crime as attack, followed by SVM by 77.5%, while Naïve Bayes by 70.4%. On the other hand, NNET NSA wrong classification of cyber-crime detection was the least, by 3.9%, naïve Bayes by 4.2% and SVM by 4.8%. By implication, our model appeared superior in the classification of cybercrime detection status (normal or anomalous).

Table 4.1: Summary of comparison of weighted average for different classifiers

Algorithm	TP Rate	FP Rate	Precision	Recall
Naive Bayes	0.704	0.042	0.512	0.592
SVM	0.775	0.048	0.420	0.650
NNET NSA	**0.857**	**0.039**	**0.857**	**1.00**

The Receiver operating characteristics (ROC) showing 90% Area under the Curve (AUC) proportion of accuracy in detection of cyber-crime using NNET NSA as shown in fig 4.4.

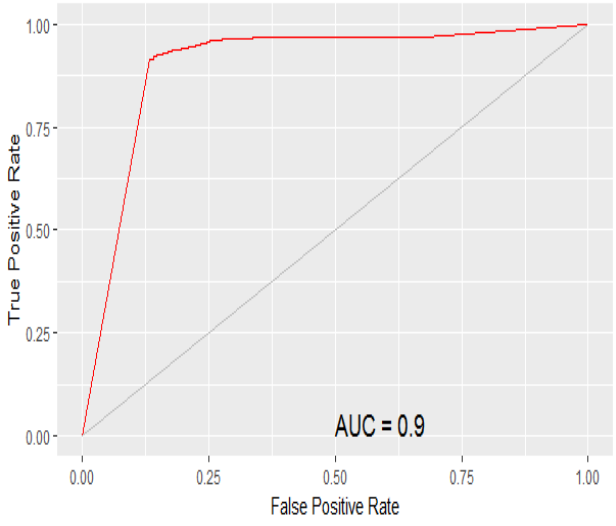

Fig5: Detection of cyber-crime using NNET NSA under AUC Curve.

5. CONCLUSION

NNET NSA yielded a higher classification accuracy of 90.1% within a lesser computational time of 15.00s, followed by Naïve Bayes 81.66% with a computation time of 100.15s, while SVM yielded an accuracy of 65.01% within a computational time of 215.81s.Naïve Bayes and SVM classifiers has been observed to be less accurate in the detection of cyber-crime attack and consume a lot of computation resources in the face of a large dataset. The result above indicates that on the R programming console NNET NSA appeared to be more efficient than other conventional algorithms in the accurate detection of cybercrime.

6. ACKNOWLEDGMENT

The authors wish to thank the Department of Computer Science, University of Ibadan for the support in this research work.

7. REFERENCES

[1] Julie Greensmith, Amanda Whitbrook, U. A. (2010). Artificial Immune Systems. nternational Series in Operations Research & Management Science, 146, 421–448, Springer Dordrecht.

[2] Aickelin, U, Bentley, P., Cayzer, S.,Kim, J., & Mcleod, J. (2003). Danger Theory : The Link between AIS and IDS ?, 147–155.

[3] Prakash, A., & Deshmukh, S. G.(2011). A multi-criteria customer allocation problem in supply chain environment : An artificial immune system with fuzzy logic controller based approach. Expert Systems With Applications, 38(4), 3199–3208. https://doi.org/10.1016/j.eswa.2010.09. 008.

[4] Dutt, I., Borah, S., & Maitra, I. (2016). Intrusion Detection System using Artificial Immune System. International Journal of Computer Applications, 44(12), 19–22.

[5] Lunt, B. M., & Ekstrom, J. J. (2008).The IT model curriculum. Proceedings of the 9th ACM SIGITE Conference on Information Technology Education - SIGITE '08,

[6] Ye, G., Wang, Y., & Sun, Q. (2019).Super Base Station Fault Detection Mechanism Based on Negative Selection Algorithm and Expert Knowledge Base. IOP Conference Series:Materials Science and Engineering, 490(07), 1-6 IOP publishing.

[7] Adeniji O.d., Olatunji O.O (2020). Zero Day Attack Prediction with Parameter Setting Using Bi Direction Recurrent Neural Network in Cyber Security.International Journal of Computer Science and Information Security (IJCSIS), Vol. 18, No. 3,pp 111-118

[8] S.D Adeniji, S Khatun, RSA Raja, MA Borhan (2008).'Design and analysis of resource management support software for multihoming in vehicle of IPv6 Network. Proceedings of the Fifth IASTED International Conference.Vol 607,issue 089.pp 13.

[9] Olushola D Adeniji, Olubukola Adigun, Omowumi O Adeyemo (2013) An intelligent spam-scammer filter mechanism using bayesian techniquesInternational Journal of Computer Science and Information Security (IJCSIS), Vol. 10, No. 3 pp 126

[10] S.D Adeniji, S Khatun, MA Borhan, RSA Raja, (2008) A design proposer on policy framework in IPV6 network.2008 IEEE International Symposium on Information Technology.Vol 4,pp 1-6

[11] Logunleko K.B., Adeniji. O.D., Logunleko A.M, (2020). A Comparative Study of Symmetric Cryptography Mechanism on DES, AES and EB64 for Information Security. International Journal of Scientific Research in Computer Science and Engineering Vol.8, Issue.1, pp.45-51.

[12] Revathi, S., & Malathi, A. (2013). ADetailed Analysis on NSL-KDD Dataset Using Various Machine Learning Techniques for Intrusion Detection. International Journal of Engineering Research and Technology (IJERT), 2(12), 1848–1853.

[13] Thabiso Peter Mpofu, D. G. V. R. R.(2014). Artificial Immune Systems: A Predictive Model for credit scoring.International Journal of Scientific Engineering Research, 5(8), 1–5.

[14] Chen, Y., Abraham, A., & Grosan, C. (2018). Cyber Security And The Evolution Of Intrusion Detection Systems. I-Manager's Journal on Future Engineering and Technology, 1(1), 74–82.

[15] Cui, L., Pi, D., & Chen, C. (2015).BIORV-NSA : Bidirectional inhibition optimization r-variable negative selection algorithm and its application. Applied Soft Computing Journal, 32,544–552.