On Development of Malware Threats Detection and Prevention using artificial Immune System and Machine Learning Algorithm

Ubochi Chinyere., Igbe C.M., Amanze B.C., Agbakwuru O.A & Agbasonu V.C
Department of Computer Science,
Faculty of Physical Sciences, Imo state University, Owerri, Nigeria
amanzebethran@yahoo.com

DOI: 10.56201/ijemt.v9.no1.2023.pg57.68

Abstract

Malare threats detection and prevention using artificial immune system and machine learning is a research work aimed at developing a system to enhance the security of the computer systems. With the rapid technological advancement, security has become a major issue due to the increase in malware activity that poses a serious threat to the security and safety of both computer systems and stakeholders. A set of malicious programming code, scripts, active content or intrusive software that is designed to destroy intended computer systems and programs is referred to as malware. According to a study, naive users are unable to distinguish between malicious and genuine applications. Thus, computer systems should be designed to detect malicious activities towards protecting the system from threats. A number of algorithms are available to detect malware activities by utilizing novel concepts including Artificial Intelligence, Machine Learning, and Deep Learning. In this study, Artificial immune system and machine learning algorithm was used for detecting and preventing malware activity. The system development utilized expert system methodology. The research work shows that adopting hybrid approaches for the development of malware detection applications provided significant advantages as 91.85% accuracy in maleware threat detection was achieved.

Keywords: *Artificial Immune System, Machine Learning, Malware, Deep Learning*

1. Introduction

In this computer age, most organizations and individuals are highly conducting transactions via the internet and this has left all day today activities highly depending on information communication technology. With this, the use of internet is growing at an exponential rate in the last decades and continues to develop in terms of dimension and complexity (Gupta, 2010). With the increase of distributed systems and data telecommunication networks, the need for automated security tools for protecting data and information became an essential requirement. One of those tools used in network security is Intrusion Detection Systems (IDSs). IDSs are software or hardware systems that automate the process of monitoring the events occurring in a computer system or a network and analyzing them for signs of security violations or unauthorized activities (Bace, 2009). For most big business and government corporations, the biggest risk of a security breach is loss of income or loss of reputation, either of which can be achieved easily by

conspicuous distributed activities such as Denial-of-Service (DoS) attacks. For organizations with more mission or life-critical data online, a DoS attack can literally put people's lives at risk. Distributed DoS (DDoS) attacks are a virulent strain of DoS activities. The difference is that there is no single source of the attack. There could be hundreds or thousands of compromised computer attackers. DDoS activities are incredibly difficult to defend against. Malware is a contraction of malicious programming codes, scripts, active content, or intrusive software that is designed to destroy intended computer systems and programs or mobile and web applications using different forms including computer viruses, worms, ransomware, rootkits, trojan, dialers, adware, spyware, keyloggers, or malicious Browser Helper Objects (BHOs) (Ahmad, 2020). Malware is the short form of malicious software or application which is not limited to computer system rather extend to the internet and related fields. Data networks have potential vulnerabilities that can be exploited by intruders. Typical attacks against these types of networks include:

1. Illicit entry, eavesdropping, unauthorized resource usage, and denial of services.
2. Node forgery or impersonation, in which legitimate cryptographic credentials are captured by an adversary, constitutes one major security threat facing data networks. The fact that communication devices are prone to be compromised and reverse engineered significantly increases the risk of such attacks in which adversaries can obtain secret keys on trusted nodes and impersonate the legitimate node.
3. The abuse of privileges by insiders to gain unauthorized access, the failure of firewall to prevent many attacks from the network, the cracking of passwords are some of the other reasons that make wireless network based system stand the risk of being attacked.

If there are inevitable attacks on the system, it will result to loss of data to unregistered users which will at times congest the network. Thus, there is need for complete protection of organizational computing resources which is deriving the attention of people towards malware threats detection and prevention.

2. AIM AND OBJECTIVES OF THE WORK

The aim of this research is to develop malware threats detection and prevention using artificial immune system and machine learning. To achieve this, the following specific objectives were set:

1. To review potential malware detection and prevention techniques and investigate the potential of Artificial Intelligence (AI)
2. To develop a system that will intelligently provide real-time detection of network systems malware attacks and prevent the attack using artificial immune system algorithm and machine learning techniques.
3. To evaluate the performance of the model in reducing network attacks in a data network when compared to existing systems.

3. Significance of the Study

There are many advantages of the new system and some of them are listed below.

1. Packets send through the system is secured and cannot be lost to attacks hence organizations will enjoy secured data transmission on the network.
2. The network security system runs with minimal human supervision and as a result reduces the cost of maintaining the network.
3. The software can monitor network intrusion and prevent possible attack.
4. The system can monitor a large number of clients on the network and trace all the activities that are carried out by a particular client.

4. Summary of the Literature Review

A lot of research was reviewed on the way to identify new threats and create secure mechanisms to counter those threats in a wireless network. From the literature's reviewed, it can be seen that in computing, a firewall is a network security system that monitors and controls the incoming and outgoing network traffic based on predetermined security rules. A firewall typically establishes a barrier between a trusted, secure internal network and another outside network, such as the Internet, that is assumed not to be secure or trusted. Firewalls impose restrictions on incoming and outgoing Network packets to and from private networks. Network firewalls filter traffic between two or more networks; they are either software appliances running on general purpose hardware, or hardware-based appliances. Table 1 summarizes some of the related studies, the techniques they adopted, the contribution and limitations of the studies.

Table 1**: Summary of Past Related Studies**

Authors	**Technique Adopted**	**Contribution**	**Limitations**
Langendoerfer, (2017)	firewall management plane (FMP)	It checks if the packet belongs to an existing connection and if the source address is already blocked	The approach cannot detect whether an incoming packet is malicious or not
Pranschke, (2009)	automated firewall rule set generation	solved time wastage which was faced when creating rule sets	useful rule sets can be neglected
Rosselti (2011)	integrate security architecture for WAN	The work try to eliminate unavailable or disrupt the connection between legitimate peers	Still leaves some loop holes for network attack
Matsunage (2012)	4-way handshake algorithm	Tries to ward-off intruders	Prevents network intrusion only

Owen (2014)	wireless intrusion detection response system	The work has 3 phase of operation in managing network security (network discovery, authentication and key generation distribution).	Prevent only session hijacking threat.
Borisov (2014)	intercepting intruders in mobile communication	The work tries to prevent adversary that are capable of doing message deletion	This approach cannot identify and prevent more attacks so it is not a robust approach
Lee (2015)	multi path approach	the approach identifies legitimate message transaction between the supplicant, authentication and authentication server	It can give false alarm
Bsufka (2016)	combination of firewalls, antivirus monitoring tools and IDS	can detect packets on the network	negative and positive false alarms can be generated and genuine packets can be denied

Though the method used is different from the existing traditional firewall, more studies are required to fully work on the challenges of traditional firewall. Method of an artificial immune system and machine learning as proposed in this research aimed at detecting and preventing any attacks or intrusions on a network as early as possible.

Research Gap Identified

Most of the work reviewed has shown that there are still some major weaknesses on the technique they adopted. The techniques they applied in their work still leaves some loop holes for network attack which is a major research gap. Thus, eliminating wireless security threats by augmenting the network with artificial immune system and machine learning technique provides strong confidentiality, integrity and replay protection for easy transmitted message in a data network.

5. Methodology

In the design of the system, Expert system methodology w adopted. Expert systems are interactive computer programs that mimic and automate the decision making and reasoning processes of human experts in solving a specific domain problem, through delivering expert advice, answering questions, and justifying their conclusions, Zaied (2005). Seflek & Carman (2010) defined expert system as a program that uses available information, heuristics, and inference to suggest solutions to problems in a particular discipline.

6. Analysis of the Existing System

The existing system is a networked setup where many communicating clients (users) exchange

files, so that one of the clients on one side can send the information, and the message could be received by the second client on the other side. The network is not secured and can be attacked by hackers. System and network technology is a key technology for a wide variety of applications. Security is crucial to networks and applications. Although, network security is a critical requirement in emerging networks, there is a significant lack of security methods that can be easily implemented. There exists a "communication gap" between the developers of security technology and developers of networks. Network design is a well-developed process that is based on the Open Systems Interface (OSI) model. So, in a network environment, tracking of user's access and activities on the network may be difficult due to lack of a highly secured system to keep track of network user's activities. There isn't a methodology to manage the complexity of security requirements. When considering network security, it must be emphasized that the whole network must be secure against malware threats. Network security does not only concern the security in the computers at each end of the communication chain. When transmitting data, the communication channel should not be vulnerable to attack. A possible hacker could target the communication channel, obtain the data, decrypt it and re-insert a false message. Securing the network is just as important as securing the computers and encrypting the message. When developing a secure network, the following need to be considered

1. Access – authorized users are provided the means to communicate to and from a particular network
2. Confidentiality–Information in the network remains private
3. Authentication–Ensure the users of the network are who they say they are
4. Integrity–Ensure the message has not been modified in transit
5. Non-repudiation– Ensure the user does not refute that he used the network.

An effective network security plan is developed with the understanding of security issues, potential attackers, needed level of security, and factors that make a network vulnerable to attack. To lessen the vulnerability of the computer to the network there are many products available. These tools are encryption, authentication mechanisms, intrusion-detection, security management and firewalls. This security approach leads to an effective and efficient design which circumvents some of the common security problems, but still malware threats is still a big challenge to computer networks.

Data Flow Diagram (DFD) of the Existing System

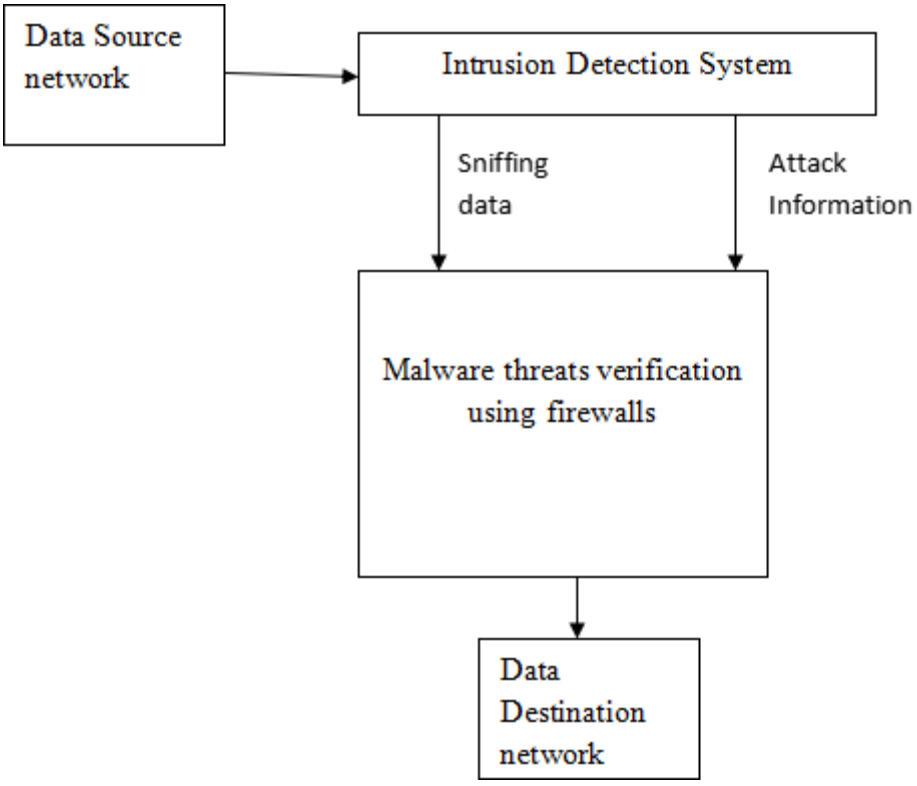

Figure 1: Data Flow Diagram of the existing system

7 Analysis of the Proposed System

Malware detection techniques are working simultaneously to detect malicious software applications. In order to improve the efficiency of malware detection techniques, improvement of existing limitations is a major fact and dynamic solutions are needed to reduce malware feature analysis time, and more sophisticated approaches should be applied to detect malicious activities. Utilizing artificial intelligence (AI) technology in the development of both malware detection and prevention needs to be increased to deal with intelligent malware that has grown in recent years. The proposed system will present an intelligent agent enhanced firewall technique using fuzzy packet filtering system, which comes with better security filtering performance. The system utilizes Fuzzy Petri Net (FPN) as a tool for modeling discrete event systems characterized by an imprecise knowledge as graphical method to describe the fuzzy logic control of packets movement through the firewall. Two levels of fuzzy will be apply to filter packets, first level led to determine the level of threat that is embedded with packets from Internet using artificial immune system, and the second level use to rearrange Access Control List (ACL) by determining the rating of acceptance and rejection of packets using machine learning algorithm. In the model, the packet filtering level relies on capturing and classifying all arrival packets depending on the information related with each packet, such as IP address, packet time and protocol type to simulate and trace the packet movement. In our approach, packet will represent by a token in FPN place, and packet's operation is exemplified by FPN transition that is responsible to move packet from one place to another. Once the packet is captured by a gate; it

moves to place where checking and matching with machine learning algorithm is done, in addition an instant copy of this packet is moved to the traffic analysis part to extract the packet's parameters (features) like number of packets coming through time period. An IDS and a network sniffer apply to detect and check whether there is any attack on the system. The IDS is used in a way that whenever any attack will launch on the network, the detection engine will send an alert to the agent. The alert could provide all the relevant information concerning the attack type, time attack happened, IP address of the attacker. Sniffing was also used to analyze the sniff output so that to the intelligent agent could check whether IDS alert is okay or not. After detecting the attack, required firewall rules are set and when it is understood that the attack ends, applied rules are canceled for increasing the performance of system. Once packets have been handled by all enabled processors, they are handed off to the detection engine. The detection engine is the meat of the artificial immune system IDS. The detection engine takes the data that comes from the processor and its plugins, and that data is checked through a set of rules (machine learning algorithm). After detection of attack, intelligent agent examines whether there is an anomaly or not. A forward chaining artificial immune system is used for anomaly detection. When the agent detects the anomaly (that the alarm is not false), prevention mechanism is triggered by the agent. Prevention mechanism consists of a rule generator, firewall and the intelligent agent. Rule generator has rule definitions for preventing various attacks. It generates firewall commands which are specified for the firewall installed on the system. After applying firewall configuration to the firewall, it should be detected whether the attack is ended or not. The attacks were cut off so as to analyze the data in the firewall logs using machine learning. This involved analyzing the logs so that see whether there any form of anomalies in the packet headers sent inform of time they were delivered. And the intelligent firewall could understand whether there is any attack regarding to the logs of the firewall and it will be prevented from attacking the computer system.

Data Flow Diagram (DFD) of the Proposed System

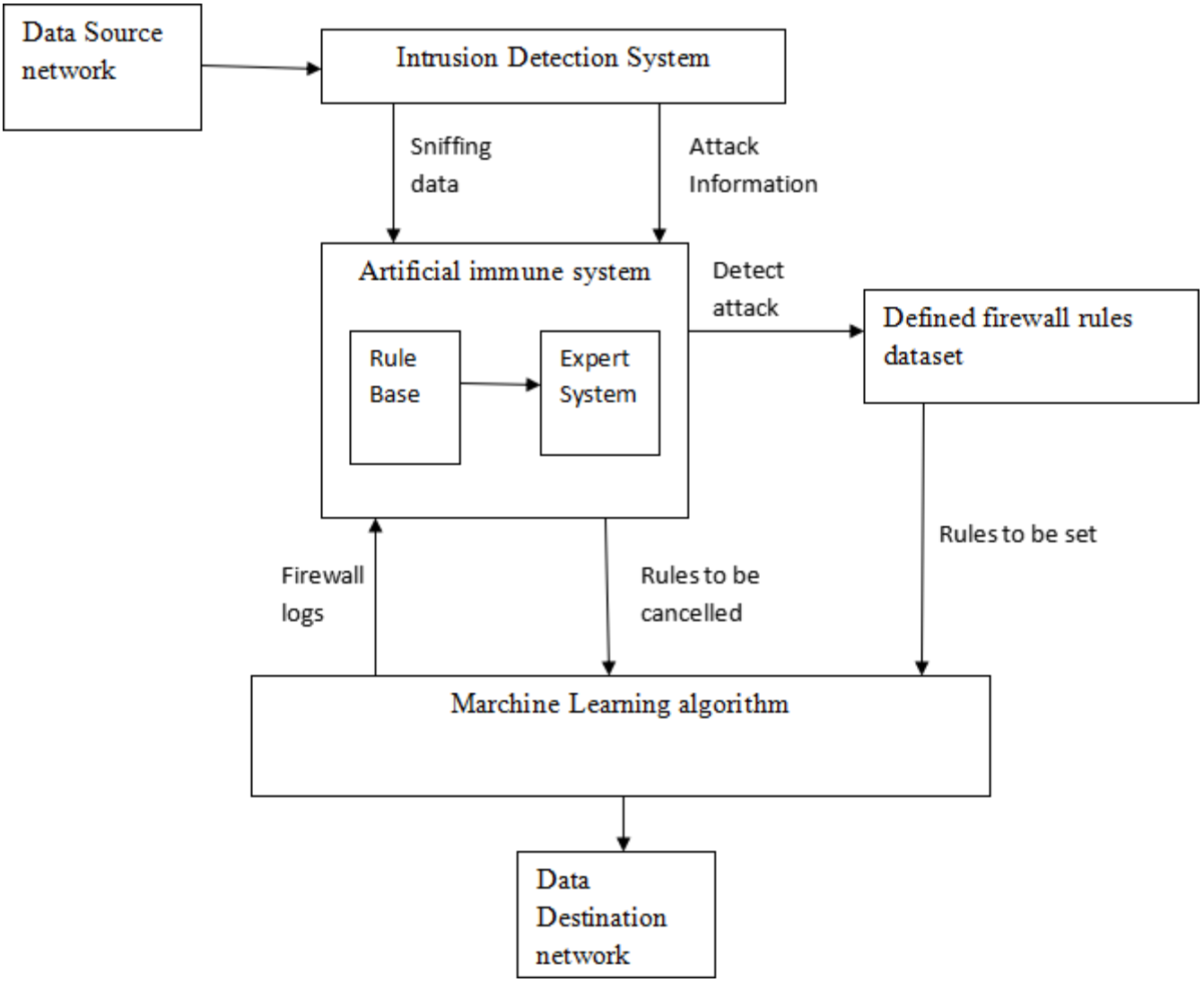

Figure 2: Data Flow Diagram of the proposed system

Performance Evaluation/ Result

Malware threats in a network increases the processing time and, in some case, degrade systems performance. The importance measure of each feature is evaluated based on the two parameters of accuracy and false positive rate. More specifically, the classification algorithm is executed with and without each feature.

This study uses some assessment metrics such as accuracy, detection rate, and false alarm rate as evaluation parameters, which are computed based on the confusion matrix.

Confusion Matrix

		Actual Class (Observation)	
		Anomaly	**Normal**
Predicted Class (Expectation)	**Anomaly**	True Positive (Correctly classified as Anomaly)	False Positive (Incorrect classified as Anomaly)
	Normal	False Negative (Incorrectly classified as Normal)	True Negative (Correctly classified as Normal)

TP: The number of correctly detected malware threats

TN: The number of harmless applications correctly recognized as harmless

FP: The number of harmless applications falsely recognized as attacks

FN: The number of attacks falsely recognized as normal.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \text{ x } 100\% \qquad 4.1$$

$$Detection\ Rate = \frac{TP}{TP + FP} \text{ x } 100\% \quad \text{Equation 1}$$

$$False\ Alarm = \frac{FP}{FP + TN} \text{ x } 100\% \quad \text{Equation 2}$$

Table 2: Malware Threat detection using genetic algorithm

True Positive (TP)	1197
False Positive (FP)	163200
False negative (FN)	7023
True negative (FN)	480010
Total No of cluster	641630

Accuracy = (1102 + 580010) / 641630 = 0.7499

The accuracy of the malware threats detection using genetic algorithm is 75%

Table 3: Malware Threat detection using Artificial Immune system and machine learning

True Positive (TP)	2304
False Positive (FP)	52293
False negative (FN)	0
True negative (FN)	587033
Total No of cluster	641630

Accuracy = (2304 + 587033) / 641630 = 0.9185

The accuracy of the malware threats detection using Artificial Immune system and machine learning is 91.85%

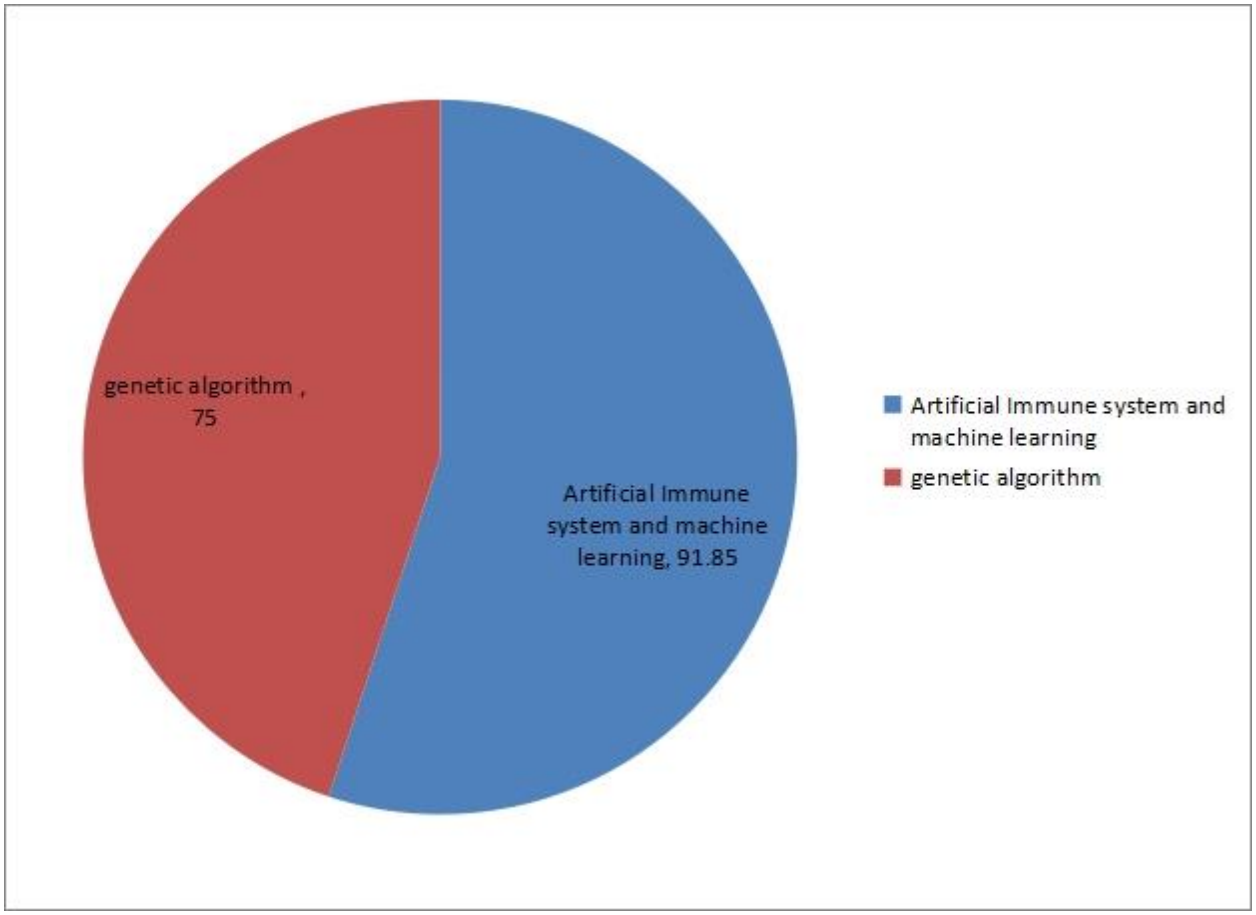

Figure 3: Performance evaluation

To evaluate our system, we focused in a major indicator of performance, which is the accuracy in detection rate and false positive rate. The evaluation of our system is done based on confusion matrix using measures 'Accuracy'. Using our approach, we have been able to achieve best results with detection rate and accuracy of 91.85% when we have used Artificial Immune system and machine learning as shown in figure 4.2. This is better than 75% accuracy obtained with genetic algorithm.

Conclusion

In this research, an overview of biologically-inspired approaches to computer security, in particular immune-inspired approaches has been examined. An intrusion detection problem, process anomaly detection, and a review of current research in this area were carried out. A malware detection and prevention system were built using immune-inspired algorithms and machine learning. The system developed was carried out using php-mysql and java script. The software developed was tested and it was able to achieve 91.85% accuracy in malware threats detection and prevention in a network.

Reference

Ahmad, M. (2020). Malware in computer systems: Problems and solutions, IJID (International

Bace, R. and Mell, P. (2009). Special Publication on Intrusion Detection Systems. Infidel

Borisov, C.M. (2014). Pattern Recognition and Matching Learning, August 2015, Publisher: IEEE, John Wiley & Sons, New Jersey, USA. Vol. 4, pg. 31-33.

Bsufka, K., Kroll-Peters, O. and Albayrak, S. (2016). Intelligent network-based early warning systems, in Lecture Notes in Computer Science, Springer, 2016, pp. 103-11.

Gupta, B., Joshi, R. and Misra, M. (2010). Distributed Denial of Service Prevention Techniques. International Journal of Computer and Electrical Engineering, 2(2), 268-276

Inc and National Institute of Standards and Technology (NIST).

Journal on Informatics for Development), 9(1).

Langendoerfer, P., Piotrowski, K., Peter, S. and Lehmann, M. (2017). Cross layer firewall interaction as a means to provide effective and efficient protection at mobile devices, Computer Communications, 7(30), 1487-1497

Lee, J., Bagheri, B. and Kao, H.-A. (2015). A Cyberr-Physical Systems architecture for Industry 4.0-based manufacturing systems," Manufacturing Letters, vol. 3, p. 18–23

Martsunage, Y.A. (2014). Secure Authentication System for Public WLAN Roaming Sept. 2014, ACM, Publisher: Association for Computing Machinery New York, Vol. 3, pg. 34-36.

Owen, H. (2014). Alternative Pair-Wise Key Exchange Protocol for Robust Dec. 2014, Publisher: IEEE, John Wiley & Sons, New Jersey, USA. Vol. 1, pg. 5-7.

Pranschke, G. C., Irwin, B. and Barnett, R. (2009). Traffic Inspection for Automated Firewall Rule Set Generation, in Information Security South Africa Conference 2009, Johannesburg

Rossetti, A.B. and Marco, K.C. (2011). Integrated Security Architecture for WLAN" June 2011, Publisher: IEEE, John Wiley & Sons, New Jersey, USA. Vol. 2, pg. 2-7.

Rossetti, A.B. and Marco, K.C. (2015). Integrated Security Architecture for WLAN Oct. 2015, Publisher: IEEE, John Wiley & Sons, New Jersey, USA. Vol. 5, pg. 9-11.

Saied, A., Overill, R. E. and Radzik, T. (2016). Detection of known and unknown DDoS attacks using Artificial Neural Networks," Neurocomputing, no. 172, pp. 385-393