JOURNAL OF CRITICAL REVIEWS

ISSN-2394-5125, Volume – 10, Issue – 7, 2023

Erratum: ABC-ANN-SVM HYBRID APPROACH TO ENHANCE CYBER SECURITY AGAINST MALWARE, DDoS ATTACKS [Journal of Critical Reviews. Vol 7, Issue 19, 2020]

***Abhishek Kajal*[1], *Sunil Kumar Nandal*[2]**

[1, 2] Department of CSE, GJUS&T, Hisar (Haryana), India

I. INTRODUCTION

In the time of COVID-19 pandemic, everything has become online, and world has become more vulnerable to cyber attacks as well. Online technology is emerging exponentially to make people more relying on computer network technology than ever earlier before [1]. Simultaneously, the total number of cyber attacks have been rapidly increased. Damage to the business continuity by these newly intrusions are increasing day by day. Computer attacks have become more complicated and lead to challenges to detect the attacks correctly [2]. The Intrusion Detection System (IDS) is being considered one of the important tools to control various types of malicious activities and threat detection what could compromise the availability, privacy or network integrity.

Today amid global lockdown, malicious attackers have considered COVID as an opportunity to launch attacks for financial gains and people are falling prey to various cyber attacks through COVID-19 related contents [3]. Inaccessibility of data may occur due to the failure of a cloud framework components, and due to any malicious activities such as distributed denial of service (DDoS), malware, spoofing etc [4]. It is important to understand the fundamentals of attacks before attempting to identify these ones. Among the most significant attacks in the contemporary environment are DDoS attacks. We can only handle DDoS attacks, in case we have enough resources. Peer systems and client-server systems, however, lack the resources necessary to overcome them. Systems therefore require a protective solution in order to identify DDoS early on [5]. DDoS attacker attempt to disrupt network in various ways. A DDoS attacker attempt to overload the network with packets before draining its resources. Mostly DDoS attacks against a single cloud user due to their resource limitations. However, it might be able to identify a DDoS attacker and stop the network's data loss. By generating fictitious packet requests, the attacker floods the communicating system's main server, preventing the main server from responding to legitimate or regular users. This causes the server to become overloaded [6]. The DDoS attack scenario where attackers get control over the main server is represented in Figure 1.

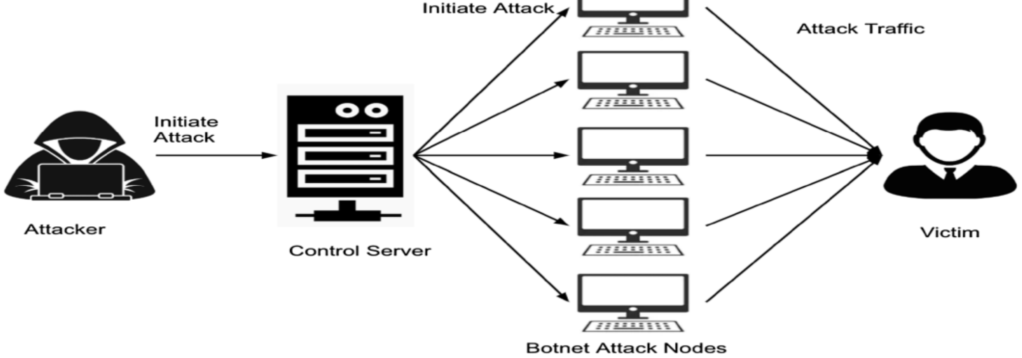

Fig. 1: DDoS Attack [7]

Second cyber attack we analyzed are Malwares which are most common attacks these days as these are designed to gain access or damage computers or networks, and the victim is still usually unaware of compromises. A common alternative description of malware is "computer virus", although there are significant differences between these types of malwares. Usually, malware will enter the device through the Internet. Although, free subscriptions of many contents, software, phishing links and every other apps the internet user download to devices or open any malicious links from the Internet without any anti-malware protection can also be the source of malware [8-9].

Third common cyber attack analyzed is Spoofing. IP spoofing is defined as one of the areas of hacker attacks when someone else's IP is used to trick the security system and penetrate an extraneous computer network. The spoofing method is used in certain targeted attacks when a hacker changes the data of the sender address in the IP packet. All these actions allow user to hide the true address of the one who makes this attack in order to receive a response packet to the address or to realize other personal goals. Most often, online intruders attack on others computer system to falsify their own IP packet headers, in particular, to change the source IP address [10-11].

The above-mentioned attacks got successfully executed due to the failure of the security mechanisms. Among the many security control methods, intrusion detection systems (IDS) are a crucial component of the defence system that shields networks from harm. IDS keeps an eye on information flow within the network and determines if an incoming packet is malicious or not [12].

In continuity in context to IDS, the main function of IDS is to alert the system if any specific harmful activity has to be taking place. IDSs can be divided into two categories: host-based intrusion detection systems (HIDS) and network-based intrusion detection systems (NIDS), depending on how the data from the incoming message is collected. IDS are divided into two categories based on how they identify attacks: knowledge-based systems and anomaly-based systems. By comparing the actions taken by a recent user to those of a typical user, the anomaly-based approach can identify the attacker. With a signature-based approach, many activities continuously monitor data traffic and compare it to a set of rules [13-14].

In this research paper, we demonstrated an IDS framework against various attacks that is taking the advantages of nature inspired algorithm with ML approaches.

The research article is organised as: Related literature work for security against different network attacks is presented in section 2. The illustration of the proposed methodology is described in the section 3. In section 4, the experimental results are illustrated by making comparison with previous work against different cyber-attacks. In last section 5, the paper conclusion defined with future prospect is defined.

II. RELATED WORK

There exist various intrusion detection systems used by a number of researchers. The detection techniques rely on the network's capabilities to identify malicious activity coming from any of the three sources: the victim, the source, and the network side. These strategies made use of statistical, machine learning, and soft computing techniques. The main objective of each approach is to identify different types of attacks and safeguard networks from unauthorized or malicious users. (He et al. 2016) [15]. This section presents the existing work performed to safeguard cloud networks against DDoS attacks. Zargar et al. (2013) suggested the Hidden Semi Markov model to detect DDoS attacks. Simultaneously, a security method based on information theory to detect any network threat is also used, that is extremely complex to get implemented [16]. This problem can be resolved by using entropy approach, which was proposed by Zhou et al. (2010) that detects threat in an accurate way [17]. This utilized the aggregation approach that will be displayed in all possible ways. The author (Yu et al. 2013) has split the data traffic into small packets and to control data traffic data features such as destination / source IP address is used. But this method fails for CSP’s that delivered services at large scale [18]. The covariance matrix has been used by (Erhan et al. 2016) that detect DDoS threat with higher efficiency, which does not rely on the assumption of normal and attacker node of the network packet distribution, but this will check the properties of entire incoming data packets to obtain accurate results [19]. Lima Filho et al. (2019) demonstrated machine learning based Denial of Service detection approach in addition to signature based approach for the extraction of features. Firstly, the signature data for DDoS and normal get extracted so as to pass to ML algorithms for the training purpose. At the end, the performance of the model has been evaluated on the basis of various parameters such as FAR, precision, sampling rate and accuracy (SR) [20]. Yadav et al. (2019) have demonstrated a k-means clustering method, based on weighted fuzzy approach in integration to neural network for the detection of malwares in cloud network with high detection rate. The designed malware detection system provides better results in terms of parameters as precision, F-measure and recall as 92.45%, 58.47%, 75.48 % respectively [21]. Samita Ranveer et al. (2015) have presented a hybrid approach for malware based on SVM classifier so as malware detection system can be strengthen to counter with different forms of malwares while keeping low false alarms and high accuracy [22]. Gu et al. (2018) have designed malware detection system by analysing statistical features from the malware family. The extraction of appropriate features minimizes the false alarm rate and hence improves the detection accuracy of the android system [23]. Sumaiya Thaseen Ikram et al. (2017) presented a new model of multi class SVM for intrusion detection system so as to reduce the training and testing time, while surging the individual classification accuracy for various cyber attacks as DDoS, malware. The experimental results of this approach reflect reduced false alarm rate and better detection rate for intrusion of different network attacks on NSL-KDD dataset, enhanced version of KDDCup 1999 dataset [24]. Watson et al. (2015) have designed a real time detection of malwares in the network. The model has used support Vector Machine (SVM) for automatic detection of DoS attack with an accuracy of 90 % [25].

III. PROPOSED WORK

In this article, enhanced security mechanism is undertaken that amalgamate nature inspired algorithms with ML algorithms. The proposed work flow (Kajal et al., 2020) [26] divided in the given sections.

i. The first section of methodology covers the feature pre-processing using the Genetic Algorithm to provide a criterion for selection of the features from the provided training datasets. It falls into two classes: normal and malicious.

ii. In the second section of methodology, wavelet transform is used to split the data into four categories, which are then processed using an artificial bee colony method to remove irrelevant features. Further, an 80-20 dataset ratio is used to separate it into training and classification sets.

iii. At the end, support vector machine and artificial neural network classifiers implemented. This way, research has been done to identify three different types of attacks: DDoS, malware, and spoofing. Figure 2 shows the framework of our security enhancement technique against cyberattacks [26].

iv. Upload dataset

Firstly, the dataset is gathered from three different sources as provided in Table 1. The data is divided into four classes: DDoS, malware, spoofing, and normal. The KDD Cup 99 dataset contains the DDoS dataset. This dataset is employed in both the validation and training phases. There are millions of records in the dataset, and each record has many relevant features. The information is available in various formats and considered dataset for identifying network intrusions. The characteristics of the dataset are useful in classifying DDoS attacked nodes.

The purpose of this dataset is to detect the probability of various malwares as well on different attributes of computers. The remote data that contains these attributes and machine infections are generated by compiling the threat reports received by the Microsoft endpoint protection solution Windows Defender. The rows in this dataset correspond to the machine, and are identified uniquely by the classification algorithm. The dataset also contains has-detection as ground truth values, which are responsible to the malware detection. The spoofing dataset is collected from the link given in table. The data contains the information related to the NCAA tournament. The data is related to sport and is of approx. of 35.8 MB size that contains around 9 tables and 201,555 rows with 106 columns.

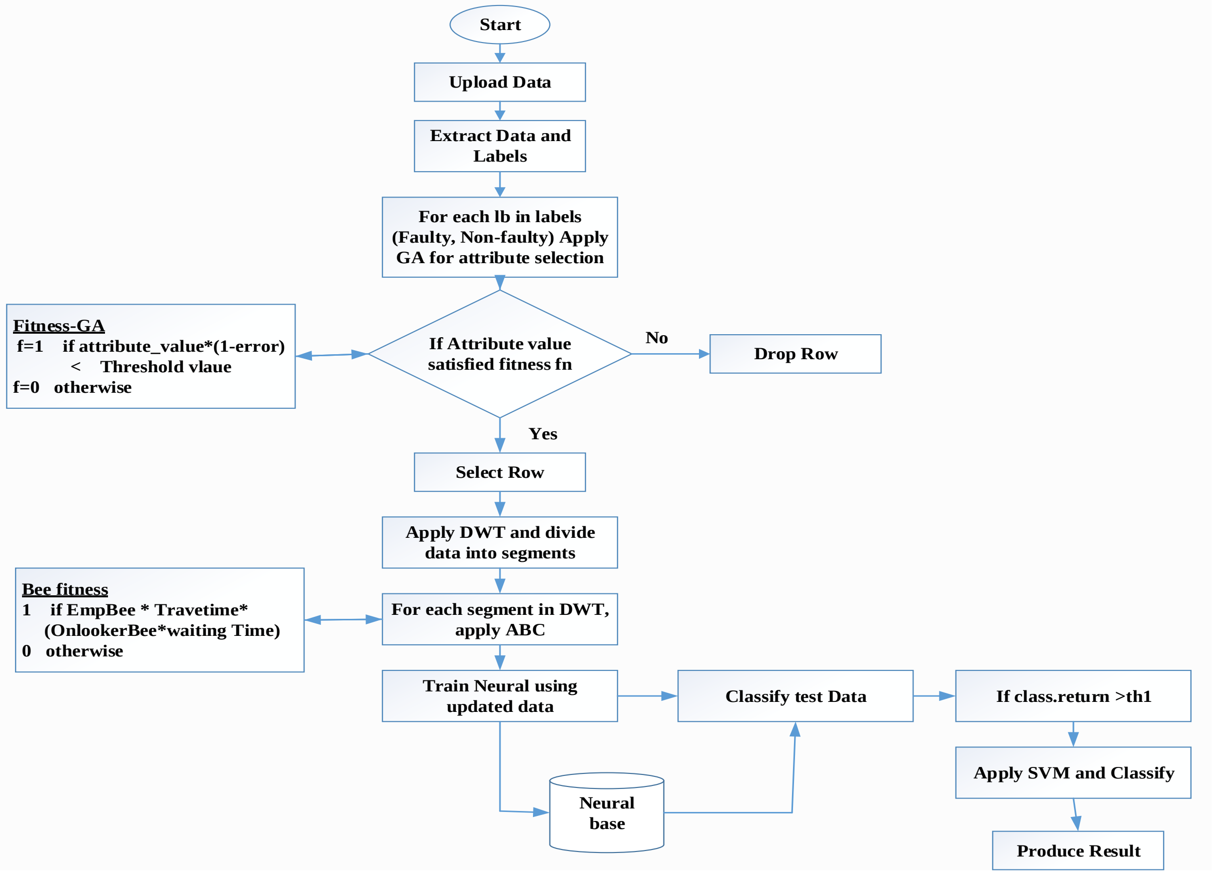
Start
Upload Data
Extract Data and Labels
For each lb in labels (Faulty, Non-faulty) Apply GA for attribute selection
Fitness-GA
f=1 if attribute_value*(1-error) < Threshold vlaue
f=0 otherwise
If Attribute value satisfied fitness fn
No
Drop Row
Yes
Select Row
Apply DWT and divide data into segments
Bee fitness
1 if EmpBee * Travetime* (OnlookerBee*waiting Time)
0 otherwise
For each segment in DWT, apply ABC
Train Neural using updated data
Classify test Data
If class.return >th1
Apply SVM and Classify
Neural base
Produce Result

Figure 2 Proposed work flow [26]

Table 1 Data set used for simulation

Category	Data Link
DDoS	KDD Cup99
Malware	https://www.kaggle.com/c/microsoft-malware-prediction/data
Spoofing	https://www.kaggle.com/c/march-machine-learning-mania-2014/data?select=regular_season_results.csv

A. Apply GA for feature row selection

Presently, most of the data is available in an unstructured format, that needs to be refined. Genetic Algorithm as a feature selection method may be used for this. ML is the most popular technique used for the selection of appropriate feature from large dataset. It filters irrelevant feature and opt out most appropriate features to improve the efficiency of ML algorithms (Zhao et al., 2010) [27].

In machine learning, the process of feature selection and construction from the large dataset is quite difficult and time-consuming. To generate features from the original data, these attributes are summarized, combined, or divided (Salguero et al. 2018) [28]. In view of the processing cost, it isn't possible to do an all-encompassing search to look out for the most relevant features. Feature selection has turned a significant problem in machine learning, which is very crucial for various applications such as regression and classification. In this, ABC algorithm is coupled with Discrete Wavelet Transformation for feature selection. To increase accuracy, ineffective features that tend to reduce learning need to be eliminated, to further decrease computational complexity (Jović et al., 2015) [29]. GA is a well-known evolutionary method that randomizes the generation of character strings. By introducing mutation, crossover like evolutionary operators, the best string among a population will be selected for most optimal solution (Hsieh et al. 2018) [30]. Genetic Algorithm is a computational

tool that is used in a variety of applications. It is very effective to solve complex problem and provide an optimal solution, especially when each objective function is varied and having numerous of local optima point (Chibaet al., 2019) [31]. The process of Genetic Algorithm begins with the generation of random population of individuals i.e. chromosomes, these are defined by a probability distribution. The probability distribution is typically uniform, updates the population at a stage called generation. Multiple individuals are selected after each iteration on the basis of a fitness function, that creates a generation with new individual's population.
GA receives about 70% of the entire data set and uses equation (1) to choose each row in the dataset.

F_s= 1 $\quad if\ (1-e)\times fs\ \times ft$
0 $\quad Otherwise$ (1)

Among three distinct attack datasets, the features are extracted to be used as feature vectors, where 0 and 1 respectively denote the attacker and non attacker. By choosing the best-optimized features to create a feature set, Genetic Algorithm is applied to the CSV dataset in order to reduce the feature set size. After optimization, the feature set is fed into DWT as input data.

C. *Apply Discrete Wavelet Transform (DWT) in addition to Artificial Bee Colony (ABC)*

Discrete Wavelet Transform technique is most effective for the processing data. Its main function is to separate and decompose data into distinct components, particularly in the frequency domain. One dimensional (1D) application of DWT splits input data into two components (L, H). LPF (Low pass filter) and HPF (High pass filter) can be used for this. The data is split into four components for two-dimensional DWT (2-D DWT) (LL, LH, HH, HL). In this work, 2D DWT is utilized to divide the selected data using GA into manageable chunks. Later, to get the optimal dataset, apply ABC to the segmented data. Here, ABC is opted for the selection of optimal feature dataset or for dataset optimization. Karaboga (2011) demonstrated the use of ABC to describe forage behaviour of bees. The scout, employees, and onlooker bees make up the majority of the bees group. Bees which search for food sources are called employee bees, and those that remain in the dance area to wait for the greatest food are called onlooker bees [32]. Scouts are tasked with conducting random searches to locate new food sources. This algorithm has onlookers at the bottom of the algorithm and artificial bees at the top. The location of the food source and the quantity of nectar from it yield a probabilistic solution to the problem, with the latter being influenced by the nectar quality of the linked solution.

This approach yields more efficient results since it is compatible with both local and global searches [32]. Equation (2) hereby provides the designed fitness function for ABC.

$$fit = \begin{Bmatrix} 1\ if\ Empbee \times Travel\ time \times (Onlooker\ Bee \times Waiting\ Time \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad Otherwise \end{Bmatrix} \qquad (2)$$

Equation (3) gives the likelihood that the onlooker bee will choose the optimal food source.

$$p_b = \frac{fit_b}{\sum fit_n} \qquad (3)$$

Where n is the population size
To determine new solution (V_{ij}), in neighbourhood of the old solution (u_{ij}), equation (4) can be used.

$$V_{ij} = u_{ij}+\varphi_{ij}(u_{ij} - u_{kj}) \qquad (4)$$

Where
i,j,k→randomly selected indexes
φ_{ij}→ random function having values [-1,1].
The ANN is trained by linking the optimized values with its weight and bias function after obtaining optimized values.

D. Artificial Neural Network (Training and Classification)

ANN (Artificial Neural Network) is a mathematical technique that works in the same manner as the human brain. There are two major category of machine learning techniques: supervised and unsupervised methods. Figure 3 represents the overall ANN structure. Numerous virtual machines (VMs) are deployed in the cloud to provide services to its consumers. In the neural network, every machine may be used to simulate multiple nodes, so as multiple VMs can perfectly be supervised by ANN. The below shown process reflects the ANN working (Pandeeswari and Ganesh, 2016) [33].

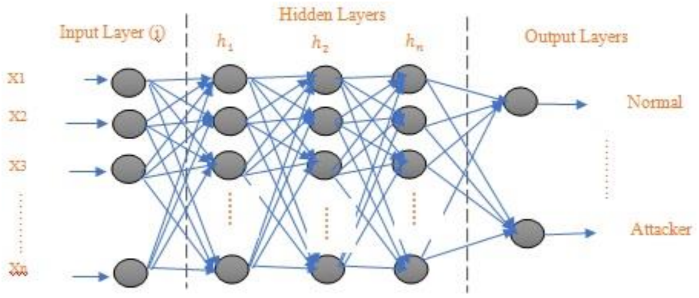

Fig. 3 ANN Structure

As represented in the fig 3, {X=X1, X2, X3 …………Xn} are the inputs present at the ANN model's input layers. Weight function Theta (θ) is used to modify weight value by hidden layer. Apply SVM to both train and test the data if the classification rate is less than 2%.

After the aforementioned steps, ANNs are trained for the normal nodes and attacker nodes and then stored in a database. Support Vector Machine should be used while evaluating the proposed work's performance if the ANN structure's classification rate is less than 2%. The dual machine learning technique aids in improving the IDS system's intended categorization rate. In Fig 2, the overall workflow is displayed. The next part provides a description of how SVM operates.

D. *Support Vector Machine (SVM)*

During the design phase, the model is used to run the test set using the training data, and it is trained with high precision by applying the model's calculation error to the training and the test inputs. The model is prepared to differentiate between an attacker and a normal node after it has been designed and appropriately determined by input training and testing.

If model doesn't work expectedly, the design process needs to be corrected. The foundation of SVM data classification is mainly based on the linearity. The lines selection with more consistent margins can be explored with linear division data. Typically, the equation to determine the optimum line of data is solved using the QP method, also referred as method of solving the problem (Sakr et al. 2019) [34].

Finding the best hyper plane to divide the data into two distinct groups is the most crucial step in this process. SVM is a classifier that creates this kind of separating line by receiving data as input. Equation (5) represents the vector that represents the line, which is evaluated as V.

$$\left||V|\right| = \sqrt{V_n^2 + V_2^2 + \ \ldots\ldots + V_n^2} \qquad (5)$$

Where, $\left||V|\right|$ represent the length of the vector, $V \ = (V_1, V_2, \ldots \ldots, V_n)$ which is termed as the norm of the vector. The direction of the vector $X = (X_1, \ X_2)$ is represented by ω in the equation (6).

$$\omega = \left(\frac{X_1}{||X||}, \frac{X_2}{||X||}\right) \qquad (6)$$

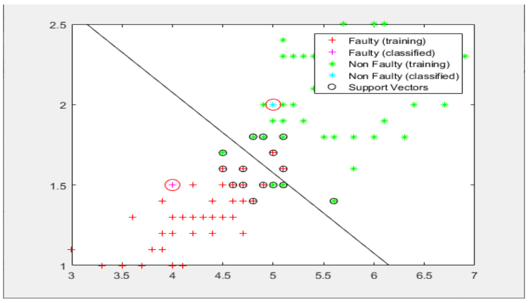

Figure 4 Classification using SVM [26]

Figure 4 represents how SVM algorithm classifies the attacks. This graph clearly reflects the classified attacker node and non-faulty nodes separately. Attacker nodes are shown by the plus signs encircle with a support vector, whereas non-faulty nodes are shown with blue stars.

REFERENCES

[1] Sharma, P., Sengupta, J. and Suri, P.K., (2019). Survey of intrusion detection techniques and architectures in cloud computing. *International Journal of High Performance Computing and Networking 13 (2):* 184-198.

[2] Khraisat, A., Gondal, I., Vamplew, P. et al. (2019). Survey of intrusion detection systems: techniques, datasets and challenges. *Cybersecurity 2 (20)*.

[3] Lallie, H.S., Shepherd, L.A., Nurse, J.R.C., Erola, A., Epiphaniou, G., Maple C. and Bellekens, X. (2020). Cyber Security in the Age of COVID-19: A Timeline and Analysis of Cyber-Crime and Cyber-Attacks during the Pandemic, *(published June 21, 2020) Cryptography and Security, Cornell University*. https://arxiv.org/abs/2006.11929

[4] Patel, A., Taghavi, M., Bakhtiyari, K. and Celestino, J. (2013). An intrusion detection and prevention system: A systematic review. *Journal of network and computer applications* 36 (1): 25-41.

[5] Preeti, D. and Kaur, A. (2016). Mitigation of DDoS attacks in cloud computing. *In 2016 5th International Conference on Wireless Networks and Embedded Systems (WECON), IEEE*: 1 - 5.

[6] Somani, Gaurav, Gaur, M.S., Sanghi, D., Conti, M. and Buyya, R. (2017). DDoS attacks in cloud computing: Issues, taxonomy, and future directions. *Computer Communications* 107: 30-48.

[7] Tuan, T.A., Long, H.V., Son, L.H., Kumar, R., Priyadarshani, I. and Son, N.T.K. (2020). Performance Evaluation of botnet DDoS attack detection using machine learning. *Evolutionary Intelligence 13:* 283–294.

[8] Hatem, Salam, S. and El-Khouly, M.M. (2014). Malware detection in cloud computing. *Int J Adv Comput Sci Appl 5 (4)*: 187-192.

[9] Saeed, Imtithal A., Selamat, A. and Abuagoub, A.M. (2013). A survey on malware and malware detection systems. *International Journal of Computer Applications* 67(16).

[10] Jeyanthi, N., and Iyengar, N.C.S. (2012). Packet resonance strategy: a spoof attack detection and prevention mechanism in cloud computing environment. *International Journal of Communication Networks and Information Security* 4(3): 163.

[11] Kang, Sung, H., Son, J.H. and Hong, C.S. (2015). Defense technique against spoofing attacks using reliable ARP table. In *2015 17th Asia-Pacific Network Operations and Management Symposium (APNOMS)*: 592-595. IEEE.

[12] Dhage, Sudhir N., and Meshram, B.B. (2012). Intrusion detection system in cloud computing environment. *International Journal of Cloud Computing* 1, *2 (3)*: 261-282.

[13] Deshpande, Prachi, Sharma, S.C., Peddoju, S.K. and Junaid, S. (2018). HIDS: A host based intrusion detection system." *International Journal of Sys*tem Assurance Engineering and Management 9(3): 567-576.

[14] Chiba, Zouhair, Abghour, N., Moussaid, K. and Rida, M. (2019). Intelligent approach to build a Deep Neural Network based IDS for cloud environment using combination of machine learning algorithms. *Computers & Security* 86: 291-317.

[15] He, Xiaofan, Dai, H. and Ning, P. (2016). Faster learning and adaptation in security games by exploiting information asymmetry. *IEEE Transactions on Signal Processing* 64 (13): 3429-3443.

[16] Zargar, Taghavi, S., Joshi, J. and Tipper, D. (2013). A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks. *IEEE communications surveys & tutorials* 15 (4): 2046-2069.

[17] Zhou, Vincent, C., Leckie, C. and Karunasekera, S. (2010). A survey of coordinated attacks and collaborative intrusion detection. *Computers & Security* 29 *(1)*: 124-140.

[18] Yu, Shui, Tian, Y., Guo, S. and Wu, D.O. (2013). Can we beat DDoS attacks? *IEEE Transactions on Parallel and Distributed Systems*, 25 (9): 2245-2254.

[19] Erhan, Derya, Anarım, E. and Kurt, G.K. (2016). DDoS attack detection using matching pursuit algorithm. In *2016 24th Signal Processing and Communication Application Conference (SIU)*: 1081-1084. IEEE.

[20] Filho, L., de, F.S., Silveira, F.A., Junior, A.D.M.B., Vargas-Solar, G. and Silveira, L.F. (2019). Smart detection: an online approach for DoS/DDoS attack detection using machine learning. *Security and Communication Networks* 2019.

[21] Yadav, R. M. (2019) "Effective analysis of malware detection in cloud computing." *Computers & Security* 83: 14-21.

[22] Ranveer, S. and Hiray, S. (2015). SVM based effective Malware Detection System. *International Journal of Computer Science and Technologies*, *6 (4):* 3361 - 3365.

[23] Jingjing, G., Sun, B., Du, X., Wang, J., Zhuang, Y. and Wang, Z. (2018). Consortium blockchain-based malware detection in mobile devices. *IEEE Access* 6: 12118-12128.

[24] Ikram, S. T. and Cherukuri, A. K. (2017). Intrusion Detection Model using fusion of chi-square feature selection and multi class SVM. *Journal of King Saud University – Computer and Information Sciences 29:* 462-472.

[25] Watson, Michael R., Marnerides, A.K., Mauthe, A. and Hutchison, D. (2015). Malware detection in cloud computing infrastructures. *IEEE Transactions on Dependable and Secure Computing*, 13(2): 192-205.

[26] Kajal, A. and Nandal, S. K. (2020). A Hybrid Approach for Cyber Security: Improved Intrusion Detection System using ANN-SVM. *Indian Journal of Computer Science and Engineering, 11 (4):* 412 - 425.

[27] Zhao, Zheng, Morstatter, F., Sharma, S., Alelyani, S., Anand, A. and Liu, H. (2010). Advancing feature selection research. *ASU feature selection repository*: 1-28.

[28] Salguero, Alberto G., Medina, J., Delatorre, P. and Espinilla, M. (2019). Methodology for improving classification accuracy using ontology: application in the recognition of activities of daily living. *Journal of Ambient Intelligence and Humanized Computing,* 10 (6): 2125-2142.

[29] Jović, Alan, Brkić, K. and Bogunović, N. (2015). A review of feature selection methods with applications. In *2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO)*: 1200-1205.

[30] Hsieh, Yi-Chih, Lee, P.J. and You, P.S. (2019). Immune-based evolutionary algorithm for determining the optimal sequence of multiple disinfection operations. *Scientiairanica,* 26(2): 959-974.

[31] Chiba, Zouhair, Abghour, N., Moussaid, K., Omri, A.E. and Rida, M. (2019). New anomaly network intrusion detection system in cloud environment based on optimized back propagation neural network using improved genetic algorithm. *International Journal of Communication Networks and Information Security,* 11(1): 61-84.

[32] Karaboga, D. and Ozturk, C. (2011). Hybrid artificial bee colony algorithm for neural network training. In *Proceedings of 2011 IEEE Congress on Evolutionary Computation (CEC)*: 84–88.

[33] Pandeeswari, N., and Kumar, G. (2016). Anomaly detection system in cloud environment using fuzzy clustering based ANN. *Mobile Networks and Applications*, 21 (3): 494-505.

[34] Sakr, Mahmoud M., Medhat A. Tawfeeq, and Ashraf B. El-Sisi. (2019). Network intrusion detection system based PSO-SVM. *International Journal of Computer Network and Information Security,* 10 (3): 22.