International Journal of Engineering

Journal Homepage: www.ije.ir

A Hybrid Modified Grasshopper Optimization Algorithm and Genetic Algorithm to Detect and Prevent DDoS Attacks

S. Mohammadi*, M. Babagoli

Department of Industrial Engineering, K.N. Toosi University of Technology, Tehran, Iran

PAPER INFO

Paper history:
Received 14 October 2020
Received in revised form 23 December 2020
Accepted 28 February 2021

Keywords:
DDoS Detection
Cyber-security
Grasshopper Optimization Algorithm
Random Forest

ABSTRACT

Cyber security has turned into a brutal and vicious environment due to the expansion of cyber-threats and cyberbullying. Distributed Denial of Service (DDoS) is a network menace that compromises victims' resources promptly. Considering the significant role of optimization algorithms in the highly accurate and adaptive detection of network attacks, the present study has proposed Hybrid Modified Grasshopper Optimization algorithm and Genetic Algorithm (HMGOGA) to detect and prevent DDoS attacks. HMGOGA overcomes conventional GOA drawbacks like low convergence speed and getting stuck in local optimum. In this paper, the proposed algorithm is used to detect DDoS attacks through the combined nonlinear regression (NR)-sigmoid model simulation. In order to serve this purpose, initially, the most important features in the network packages are extracted using the Random Forest (RF) method. By removing 55 irrelevant features out of a total of 77, the selected ones play a key role in the proposed model's performance. To affirm the efficiency, the high correlation of the selected features was measured with Decision Tree (DT). Subsequently, the HMGOGA is trained with benchmark cost functions and another proposed cost function that enabling it to detect malicious traffic properly. The usability of the proposed model is evaluated by comparing with two benchmark functions (Sphere and Ackley function). The experimental results have proved that HMGOGA based on NR-sigmoid outperforms other implemented models and conventional GOA methods with 99.90% and 99.34% train and test accuracy, respectively

***doi**: 10.5829/ije.2021.34.04a.07*

1. INTRODUCTION

The last decades have witnessed a revolution of Internet usage in many domains like e-commerce, e-government and so on. The expansion of the Internet is accompanied by the intensification of security violation issues. Denial of Service attack (DoS) is an intimidating attack which targets servers, online resources and network bandwidth. Victim's resources such as processors, bandwidth, database, memory, etc. are occupied with packet flooding which is generated by a malicious person or bot [1]. The devastation of servers or causing interruptions in online services is considered as the principal purpose of this attack. Distributed denial of services attack (DDoS) emerged as a powerful version of DoS with the capability of inflicting more destructive damage in a shorter span of time. Typically, DoS attacks are launched using one computer and one internet connection, whereas DDoS attacks are carried out by using several compromised computers (bots) and internet connections. Figure 1 shows one type of DDoS attack with multiple bots. In this figure, masters and slaves are hired in conjunction with an attacker to generate an enormous amount of packet [2, 3].

1. 1. DDoS Classification In DDoS attacks, the malicious user hires a network of zombie computers to incapacitate a server or website. DDoS attacks are categorized into three main groups: volume based attacks, protocol attacks and application layer attacks. Volume based attacks is the most common type of the aforementioned groups. These attacks send a large amount of requests or data to the victim's server with

*Corresponding Author Institutional Email: Mohammadi@kntu.ac.ir (S. Mohammadi)

Please cite this article as: S Mohammadi, M Babagoli, A Hybrid Modified Grasshopper Optimization Algorithm and Genetic Algorithm to Detect and Prevent DDoS Attacks, International Journal of Engineering, Transactions A: Basics Vol. 34, No. 04, (2021) 811-824

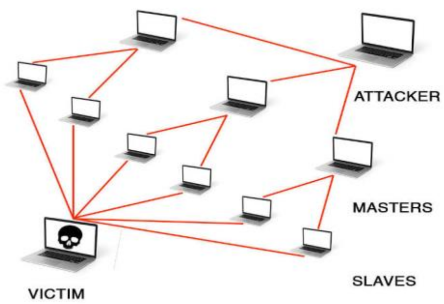

Figure 1. DDoS attack procedure

the purpose of overwhelming its bandwidth capability. Unavailability is considered as a major consequence of this type of attacks. Volume based attacks are prevalent in light of their simple amplifications; then, script kiddies can utilize this method for harming specific web services. Unlike the volume based attacks, Protocol features are abused in Protocol attacks [4]. What is employed in this type of DDoS attacks is an attempt to destruct or suspend a web resource. Indeed, intermediate communication devices (like load balancers and firewalls) are targeted to disrupt the communication of a website and its server. On the other hand, zombies (bots) are utilized in application layer attacks (or a 7-layer attack) to penetrate a specific server using the application's vulnerabilities [5]. This type of attacks requires fewer resources in comparison with the mentioned types on the grounds that it focuses on specific application packets which are sent through normal HTTP requests. Consequently, detection of application layer attacks is considered to be a laborious procedure [6]. The classification of DDoS attacks is described in depth in Figure 2.

1. 2. DDoS Prevention and Detection Despite there being a lot of DDoS detection and prevention methods, deterring such attacks effectively is far-fetched if not impossible. In fact, the mitigation of DDoS risk has been the main aim of researchers. On the other hand, tracing back to the source is impractical as a result of IP spoofing (IP address is forged), stateless nature of network and similarity to flash crowd [7]. Therefore, source attack identification in DDoS attacks is an onerous endeavor. The detection and prevention techniques are divided into 3 categories: trace back methods, entropy based detection and intrusion detection and prevention systems [8].

Trace back methods have enhanced routers and protocol capabilities to track packets and uncover the source of attack. This method is often costly and with low accuracy. Packet marketing scheme and IP trace back technique are two schemes of this method [9]. Entropy is a measure of information theory which scales randomness of packets on specific router in entropy based DDoS detection. Indeed, the changes of flow's (packets with same destination address) abnormality are measured using entropy and the alarm would be raised if the rate of entropy is large. Hence, by tracking the entropy variation, the source of package is obtained. Information distance is the next metric which is used for distinguishing DDoS attacks and flash crowd. Intrusion detection system (IDS) is used to monitor the web traffic and report any suspicious activity to the administrator and intrusion prevention system (IPS) is designed to detect and prevent the attacks together with analyzing the data flow [10]. The segmentation of DDoS detection methods is illustrated in Figure 3.

In order to prevent DDoS attacks, many researchers have proposed different methodologies which focus on detection, prevention and trace back. Nevertheless, the lack of considering the limitation of real-time problems, complexity and massive data is a critical issue in DDoS detection strategies. With the intention of solving the aforementioned problem, anomaly based detection methods are used to create a profile of the normal traffic and then, detect the unknown attacks. Machine learning techniques are used to model a reliable behavior in network domain as a reference, and then compare new

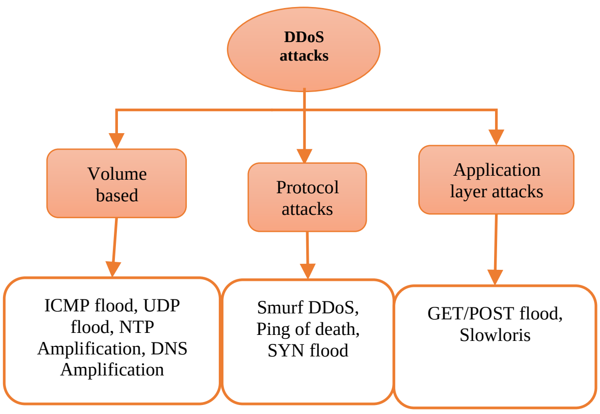

Figure 2. DDoS classification

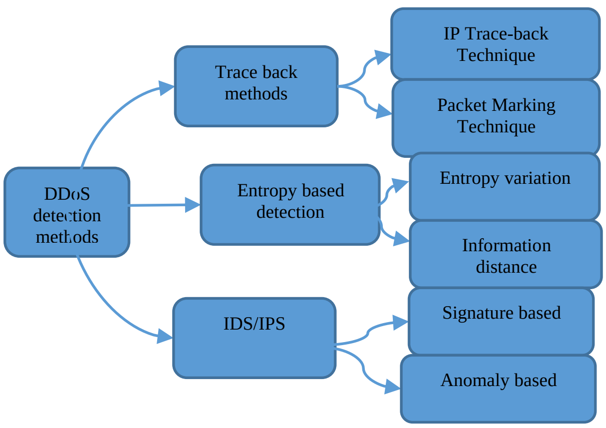

Figure 1. Classification of DDoS detection methods

ones with it. On the other hand, meta-heuristic algorithms are a nominated strategy to address the complexity and real-times issues and these algorithms can solve (NP)-hard problems [11]. additionally, some other important features like easy to use, cost-efficient and preparing important tools for both researchers and managers to solve the complex dilemma, makes these algorithms more popular [12]. Considering the no free lunch theory, there is no guarantee to one meta-heuristic algorithm outperforms in all problems. In order to reach better performance in a specific problem, there are several new meta-heuristic algorithms can be proposed or conventional algorithms can be modified or different algorithms can be combined with each other [13]. Accordingly, a combination of machine learning techniques and meta-heuristic algorithms can be used to boost the performance of detection method in terms of accuracy, speed and extendibility[14].

In this paper, the combination of machine learning and meta-heuristic algorithms is utilized to resolve the issue at hand and an efficient model is proposed to detect and prevent DDoS attacks. In order to evaluate the performance of the proposed method, two benchmark cost functions are applied into the model. An up-to-date dataset (CICIDS) is used to train and test the model. CICIDS consists of reliable and real-world samples which cover different attacks properly [15]. Ultimately, one machine learning technique: random forest (RF) and two meta-heuristic algorithms (Hybrid Modified Grasshopper Optimization algorithm and Genetic Algorithm (HMGOGA) and conventional Grasshopper optimization algorithm(GOA)) are utilized for feature selection and DDoS detection, respectively. The rest of this paper is organized as follows: section 2 is devoted to a review of related literature regarding the previous studies, section 3 touches upon the proposed DDoS detection method, section 4 describes and discusses the experimental results and finally, section 5 concludes the present research.

2. LITERATURE REVIEW

Taking into account the background of DDoS attacks, some major DDoS detection mechanisms are described in this section. The focus of this section is on machine learning and data mining approaches. Gu et al. [16] utilized the semi-supervised weighted k-mean and hybrid feature selection (SKM-HFS) method to detect DDoS attacks. In order to validate their experiments, they used three benchmark datasets and the results of their proposed mechanism were compared with one another. The feature selection performance was evaluated using TOPSIS method. As shown in their results, SKM-HFS had better performance in both time-consumption and precision. Finally, with the purpose of evaluating SKM-HFS in the real world, a real experimental environment was employed to appraise the functionality of the proposed algorithm. Like other experimented datasets, SKM-HFS has shown an acceptable performance in the real world dataset. Gharvirian et al.[6] used a perceptron neural network along with computing entropy of flow and flow initiation rate in order to detect the DDoS attack in the SDN controller. Indeed, In this research, the neural network makes improvement in the detection accuracy and false alarm rate and proves the existence of attack by investigating the 3 features of network traffic. Considering the vitality of the detection time, the proposed model used the neural network just for suspicious flows. The detection accuracy approximately reached 92% and the delay of detection obtained 23.55 seconds which is proof positive of the detection efficiency. Ghasemi et al.[17] proposed a multi-stage detection model and in each stage, they concentrated on one type of attack. They used genetic algorithm in order to select the most important features of each type of attack. In this paper, a novel chain model is proposed to detect each type of attack respectively. After one type of attack is detected, the chain model deletes specific labels from the dataset. In order to evaluate the proposed model, two benchmark datasets (NSL-KDD and KDD cup99) were used. The accuracy of average detections for two datasets were 97.5% and 98.9%, respectively. Four different classifiers are used as the fitness function for genetic algorithm, decision tree outperforms other methods in most cases. Nezhad et al. [18] have applied time series model and chaotic system to distinguish between legitimate and suspicious traffic. Two features (number of packet and number of source IP address) have been used as detection metric in every minute, and a detection accuracy of about 99% has been obtained. The Box-Cox transformation, Auto Regressive Integrated Moving Average (ARIMA) and Lyapunov were utilized for data processing, predicting and classification phases, respectively. Many DDoS detection methods based on machine learning were tried on SDN[2]s (Figure 4). Artificial neural network was employed to detect the different types of DDoS attacks [19]. ICMP flood, SYN flood, UDP flood and DNS reflection were experimented using proposed collaborating intrusion detection system (CIDS). The emulation results have proved the proficiency of ANN[3]-based CIDS in SDN. Conversely, some inherent features of SDN can be used to assist the confrontation with DDoS attacks. In this trend, SDNs advantages can be used for DDoS detection in cloud environment [20]. The methodology for DDoS defeating in SDN can use learning techniques (Machine learning/Deep learning)

[2] Software define network
[3] Artificial neural network

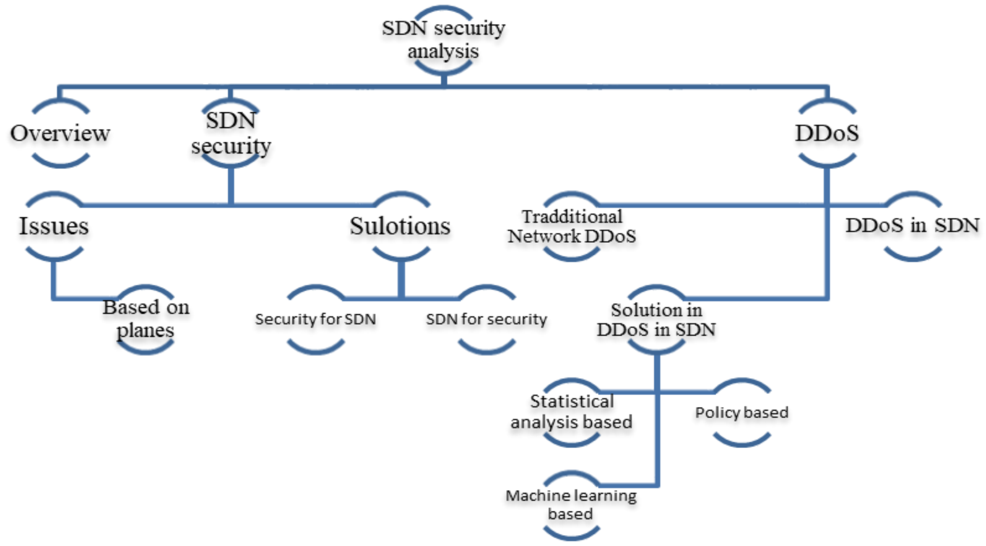

Figure 4. Research categories in SDN security domains [25]

to ameliorate detection rates and reduce the computation cost and time. Niyaz et al. [21] have proposed a network application on the basis of deep learning for multi-vector attack detection. Deep learning methods have been employed to remove irrelevant features and select the most important ones. Three implemented models were investigated for feature classification and the accuracy of 95.65% was obtained from SAE[4] (stacked sparse auto-encoders and soft-max classifier) approach. Arivudainambi et al. [22] have proposed Lion optimization algorithm [23], a new meta-heuristic algorithm, to detect DDoS attacks in SDN. The vector feature selection method has been applied to the selected dataset (NSL-KDD) to collect an appropriate feature subset and a combination of Lion Optimization Algorithm and Convolutional Neural Network has been used for training and testing. As it was demonstrated in their results, the average accuracy reached to 96%. Sreeram et al. [24] proposed a bio-inspired bat algorithm to detect HTTP flood attack in a short time frame and with high speed. The CAIDA dataset was used to select the most important feature for the proposed model. Afterwards, the selected features were used to train and test the bat algorithm. As shown in their results, they have obtained 94.8% accuracy in detecting HTTP flood attacks.

However, most of the researches mentioned above were incapable of adequately detecting new DDoS attacks at the right time. Some of the main drawbacks of the existing literature which were used as motivation for our research are as follows: lack of high accuracy accompanied by acceptable time-consumption and extendibility, difficulty in detection of unknown and zero-day DDoS attacks, lack of expansion of new methodology for detecting DDoS attacks, not using comprehensive datasets, etc.

[4] Stacked auto encoder

3. PROPOSED DDoS DETECTION MODEL

Having plenty of information in networks packet, abnormal behavior of packets can be recognized using analysis methods. Therefore, some available datasets are included in network data to provide efficient context for network security researches. NSL-KDD [26] and CICIDS [27] are the two most popular datasets that have been provided for network threats investigations. Due to antiquated data in NSL-KDD. This research has employed CICIDS in order to assess the proposed model. Some traffic features in CICDS are ineffectual, leading to degradation of learning quality, more memory consumption and an increase in computational time. Feature selection methods can properly solve these issues. In this paper, a machine learning method, RF, is used to collect more important features. Next, HMGOGA is utilized to detect DDoS attacks using the selected features. At last, a comparison of the conventional method and other research is made to evaluate the performance of our model.

3. 1. CICIDS Dataset The DDoS dataset applied in this manuscript is adopted from UNB repository [27]. The dataset consists of 77 features and one label column. The types of traffic are indicated using label column. Due to the problem of diversification in other datasets, CICIDS comprised 225,745 samples which include legitimate and attack traffic. The feature description is available by details in [15]. Table 1 provides the general information about CICIDS.

3. 2. Feature Selection Feature selection methods, a type of dimensional reduction techniques, are used to transform features into a new space with low dimensions. Indeed, the irrelevant features are eliminated from the set of features and the most important ones remain [28]. Prior to our DDoS detection method, RF is used to improve the detection throughput. RF as a popular machine learning method makes use of tree based decision making and results in an efficient performance regarding the low over-fitting, good predictive accuracy and ease of use [29]. The relevant features are selected by their impurity measures; as a matter of fact, when a tree is trained in RF, decrease of weighted impurity in a tree can be computed by each feature. Therefore, the average of each feature's impurity reduction can be used to rank the features in a forest. According to correlated features in CICIDS, the most important features have led to low impurity [30]. Continuing this process, selected features are used as feed for meta-heuristic algorithm and the procedure goes on to detect the DDoS attack. Figure 5 shows the framework of the proposed detection method.

3. 3. Proposed DDoS Detection Method In the detection phase, initially, the GOA is trained using the selected feature subset to develop an ability to detect unknown attacks. The GOA is a new meta-heuristic algorithm that is inspired by the behavior of grasshopper swarms while finding food and moving toward the source of food. The mathematical model of this algorithm is used for optimization problems [31]. The GOA is based upon swarm intelligence and population based categories. The merits of GOA algorithms are proved using several test functions and it is outperformed in cases of productivity from exploration to exploitation, randomness quality, search space coverage, scape from local minimum and fast convergence to optimum solution [32]. The mathematical model of GOA is described below. The position of each grasshopper is obtained using Equation (1).

$$X_i(t+1) = S_i(t) + G_i(t) + A_i(t) \tag{1}$$

where, S, G and A denote the Social Interaction (SI), gravity force and effect of wind flow, respectively.

The social interaction is the main parameter of GOA and plays a pivotal role in problem optimization. Social interaction is defined as follows:

$$S(i) = \sum_{j=1(j \neq i)}^{nPop} s(d_{ij})\hat{d}_{ij} \tag{2}$$

where, d_{ij} denotes the Euclidean distance between the i th and j -th grasshopper and $s(d_{ij})$ is a social force function that is based on attraction and repulsion forces. Hence, the effect of grasshoppers on each other is measured using this function.

$$s(d) = fe^{-\frac{d}{l}} - e^{-d} \tag{3}$$

where, f is gravity intensity and l is gravity length scale. By rewriting the main equation:

$$X_i = \sum_{j=1(j \neq i)}^{nPop} s(|x(i) - x(j)|)\frac{x(i) - x(j)}{d_{ij}} - ge_g - ue_w \tag{4}$$

where, e_g, e_w and u denote the unit vector across the direction to the center of the earth, unit vector across

TABLE 1. Details of CICIDS

	# of rows	# of column	# of legitimate traffic	# of attack traffic
CICIDS	225745	78	97718	128027

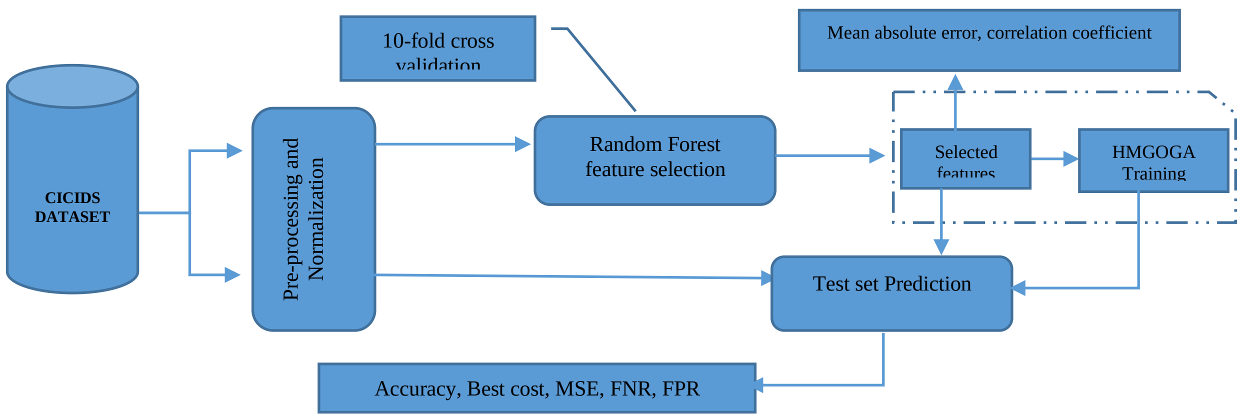

Figure 5. Proposed model framework

wind blow direction and fix drift, respectively. The proposed equation is not usable for optimization problems due to the weakness of exploration and exploitation in finding the optimal solution, thus utilizing the modified equation in this paper as follows:

$$X_i^d(t+1)=c\left\{\sum_{j=1(i\neq j)}^{nPop} c\frac{ub_d-lb_d}{2}s\left(\left|x_i(t)-x_j(t)\right|\right)\frac{x_i(t)-x_j(t)}{d_{ij}}\right\}+\hat{T}_d(t) \tag{5}$$

Where, ub and lb are the upper bound and lower bound in d-th dimension of Equation (6) and $\hat{T}(t)$ denotes the best solution that has been found so far.

GOA is useful for solving many complex global optimization problems. Nevertheless, there are some drawbacks in the conventional GOA like low convergence speed and stuck in the local minima [33]. Due to the complexity of our search space, the position of each grasshopper must update more accurately considering the whole search space. In order to reduce the time of finding the optimal solution and increase the convergence speed of GOA, a new SI strategy has been introduced in this paper. In the conventional GOA, the social interaction for each grasshopper can be obtained using the distance between one grasshopper and others. Indeed, in each iteration, the specific grasshopper can be affected by both far and close grasshoppers equally (Equations (2) and (3)). Consequently, improper effects of far grasshoppers cause an increase in computing time and algorithms iterations in order to find the optimal solution. In this paper, a novel strategy is introduced for SI which moderates grasshopper effects. Social interaction force is calculated for each grasshopper using just the nearest grasshoppers, not far ones. Indeed, by organizing the position of grasshoppers, the speed and power of finding global optimum is increased; however, sometimes this algorithms may gets stuck in local optima for complex optimization problem due to its weak diversity [34].

$$S(i)=\sum_{j=1(j\neq i)}^{nNearest} s(d_{ij})\hat{d}_{ij} \tag{6}$$

Unbalanced exploration and exploitation is another weakness of the original GOA that can lead into falling in a local optimum trap [35]. To overcome this obstacle, in this research, two genetic algorithm principles, crossover and mutation, are added to the GOA. Crossover and mutation operators in GA, work for diversification and intensification phases and one of the main characteristics of the GA algorithms is the behavior of operators that operates by chance. Although this characteristic is considered as negative point of GA, makes our model more powerful in the exploration phase. This proposed algorithm- called Hybrid Modified Grasshopper Optimization and Genetic Algorithm (HMGOGA)- is considered to be an extension of MGOA which enhances the exploration and exploitation power of the algorithm for the purpose of avoiding local minimums. In further detail, in each iteration, after the grasshopper position has been updated (Equation (5)), parents are selected from a new grasshopper population and offspring created by exchanging genes. Parent selection is randomly uses one of the following three methods in each iteration: Roulette Wheel, Random and tournament selection. Subsequently, binary crossover is applied to the selected parents and offspring can be created. Before adding the offspring to grasshopper population, in the mutation phase, some of the grasshopper genes are flipped randomly. The exploration and exploitation capabilities of modified GOA (MGOA) are improved using crossover and mutation, respectively. In order to fair operation of exploration and exploitation, the c parameter in Equation (5) is decreased by increasing iteration. The detailed Pseudo code for the proposed method can be described as follows and the flowchart of proposed algorithms is illustrated in Figure 6.

```
Start
Initialized parameters and population
for i=1:MaxIteration
        - update all grasshopper position (Eq. 5). Social interaction for each grasshopper is calculated just by closer grasshoppers.
        - evaluate population using cost functions
        - generate random number between 1:3 to determine parents' selection strategy
        - apply crossover and create offspring
        - apply mutation for random grasshoppers' genes.
        - evaluate offspring using
        - concatenate created offspring and grasshopper population and select the best population.
End
Finish
```

In order to find near-optimal solution in meta-heuristic algorithms, parameter tuning is a major concern of researchers for improving efficiency and capability of algorithms. Parameter tuning provides more flexibility and robustness in problem solving and it requires careful initialization. Indeed, the parameter tunning is highly related to the complexity of the problems but many researchers propose an optimal value for key parameters of the algorithms [36]. In this research, after using trial and error method for finding best value in algorithm setting, the researcher's proposition is used. For instance, in order to fair usage of exploration and exploitation proportional in Equation (5), the c parameter is calculated as follows [35]:

$$c=c_{\max}-currentIt*\frac{c_{\max}-c_{\min}}{\max It} \tag{7}$$

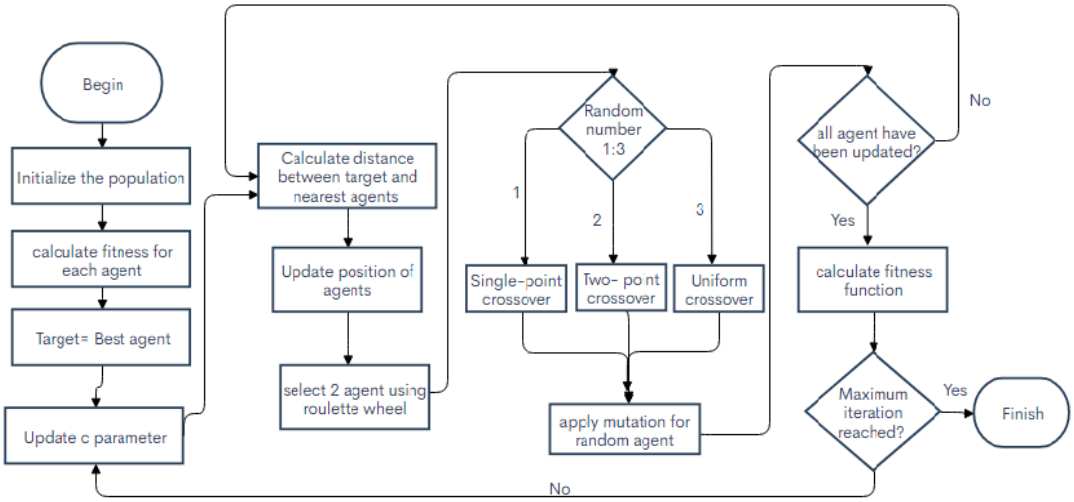

Figure 6. Flowchart of proposed model

where, $c_{\max}$ and $c_{\min}$ is maximum and minimum value, $currentIt$ is a current iteration and $\max It$ indicates the maximum iteration. According to Equation (7) the c parameters reduce in each iteration. In fact, the c parameter is updated to reduce exploration and increase exploitation ($c_{\max}$ and $c_{\min}$ are considered 1 and 1e-4, respectively).

In order to prove the efficiency of the proposed model, several benchmark functions are used as cost functions for GOA. In fact, the optimum coefficients of cost functions are calculated using the meta-heuristic algorithm. Three benchmark functions are applied so that the model performance can be figured out in various conditions such as Sphere, Ackley function [37, 38] and non-linear regression [39]. The sphere function is a simple continuous, convex and unimodal function which is widely used for optimization problems [40]. Ackley function is utilized as a more sophisticated function in the proposed model. Ackley function was first applied to genetic hill-climbing [38]. Ackley function is a non-convex function which is used for testing the optimization algorithms. Nonlinear regression is a form of linear regression analysis in which the relation between dependent and independent variables are nonlinear. Regression analysis mainly aims to model the observational data and find the relationship between responsible variables (y) and predictors (x). The relation between x and y is investigated using coefficients and the optimal values of the coefficients are obtained through using meta-heuristic algorithms. The implemented equations are shown in Table 2.

Where, α is the parameter that must be optimized, β is a random number between [-1, 1], x denotes the input vector which consists of the selected features, and n is the number of the selected features.

After the training phase, GOA predicts the label of test data by using cost functions and then, the accuracy of prediction is evaluated using Mean Square Error(MSE). The MSE formulation for both train and test phases is shown in Equation (8).

TABLE 2. Implemented benchmark cost function

Name	Cost functions	Equation number
Sphere	$y=\sum_{i=1}^{n}\alpha_i x_i$	(9)
Ackley function	$y=-20\times exp(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n}a_i x_i^{2}})-exp(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi a_i x_i))+20+exp(1)$	(10)
Combined Nonlinear regression-Sigmoid function	$y=\dfrac{1}{1+e^{-(\sum_{i=1}^{N}\alpha_i x_i+\sum_{j=1}^{N}\sum_{k=j+1}^{N}\alpha_{jk}x_j x_k+\beta)}}$	(11)

$$MSE = \frac{\sum_{z=1}^{N}(class\_label(z) - f(s))^2}{N} \quad (8)$$

where, $f(s)$ shows the desire output, N represent the number of rows in the dataset and $class\_label(z)$ denotes the real class of each packet.

4. SIMULATION RESULT

In this section, HMGOGA algorithm is applied to several cost functions and the results have been thoroughly compared. In order to evaluate the proposed model, some credible research projects have been compared and the efficient application of our proposed method is investigated using DDoS detection. Firstly, data are cleaned and the null features are removed from CICIDS dataset in the pre-processing phase. Next, the data will be normalized for the purpose of homogenization of features effect (Equation (12)).

$$x' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (12)$$

where, $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of each feature, respectively. After normalization, 70% of data is randomly selected for training and the other 30% is kept and used for testing evaluation. In order to eliminate redundant features, improve detection accuracy and reduce the computational cost and required storage, RF is utilized as a feature selection technique. As it can be deducted from the results, 20 features are efficiently selected among 77 features and the performance of the selected ones is validated by Decision Tree (DT). Figure 7 shows the ranking of the selected features on the basis of their ranking merit and the details of these selected features are described in Table 3.

TABLE 1. Details of selected features

Number	Name	Merit
feature 44	ACK Flag Count	0.083998
feature 0	Destination Port	0.078377
feature 45	URG Flag Count	0.061353
feature 10	Bwd Packet Length Max	0.055670
feature 12	Bwd Packet Length Mean	0.054299
feature 50	Avg Bwd Segment Size	0.047392
feature 47	Down/Up Ratio	0.039466
feature 48	Average Packet Size	0.037646
feature 13	Bwd Packet Length Std	0.036543
feature 38	Packet Length Std	0.033712
feature 37	Packet Length Mean	0.033700
feature 36	Max Packet Length	0.027459
feature 8	Fwd Packet Length Mean	0.023853
feature 59	min_seg_size_forward	0.023748
feature 6	Fwd Packet Length Max	0.018635
feature 43	PSH Flag Count	0.017955
feature 49	Avg Fwd Segment Size	0.017868
feature 22	Fwd IAT Std	0.017700
feature 39	Packet Length Variance	0.017284

As shown in the results, the most valuable features are extracted efficiently and about 75% of the irrelevant features are removed from the subset. For assessing the feature subset using DT classifier, two measures are used: Mean Absolute Error (MAE) and Correlation Coefficient.

The distance between two variables is measured using MAE that is investigated in this phase to calculate the average absolute difference between the prediction and true class label values. The strength of relation between two variables is obtained by Correlation Coefficient metric. Indeed, the dependence of features to the labeled class is defined using correlation coefficient. The competency of feature subset is proved by high correlation and low MAE for which 96.84% and 3% were obtained, respectively.

Henceforth, the HMGOGA algorithm is qualified to detect DDoS attack using the most important features. In order to strike high performance strategy, different benchmark test functions are utilized and the parameters of functions are optimized to decrease the MSE and increase the detection accuracy. Primarily, like other meta-heuristic algorithms, a random population value between [-1, 1] is generated as coefficients of target functions. In each iteration the powerful particles (grasshoppers) are maintained and the weakest ones are eliminated. The strength of particles is defined using MSE. As a matter of fact, each row of population is multiplied into the target function using training dataset and the population is changed in each iteration according to HMGOGA procedure.

Finally, the most eligible particle is considered as an elected coefficient. In this step, 70% of data is selected randomly for training and the other 30% is considered as test data. Figure 8 (b, d, f) demonstrate the training phase of HMGOGA algorithm with different cost functions. As it can be observed in Figure 8, the downward trend of the MSE indicates the successful process of training. The performance of the proposed model is checked by predicting the precision of test data

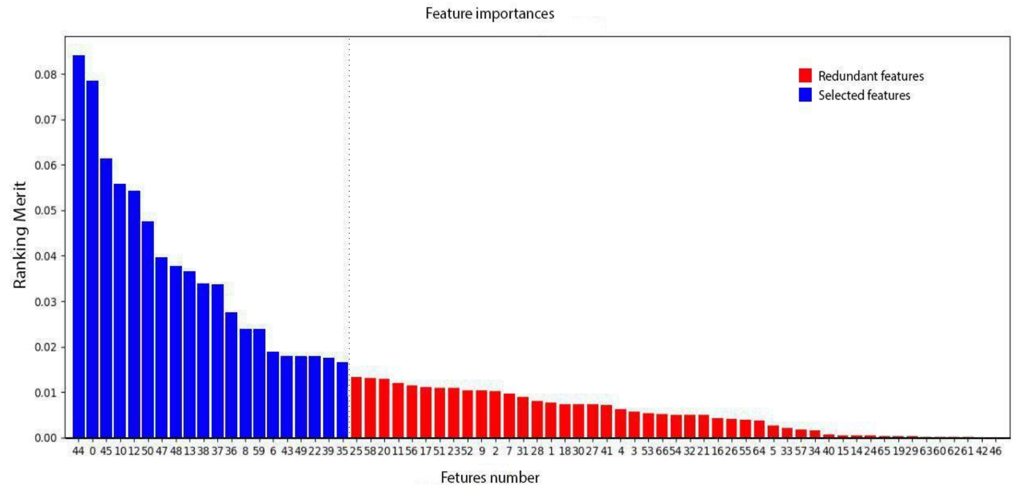

Figure 7. Features ranking

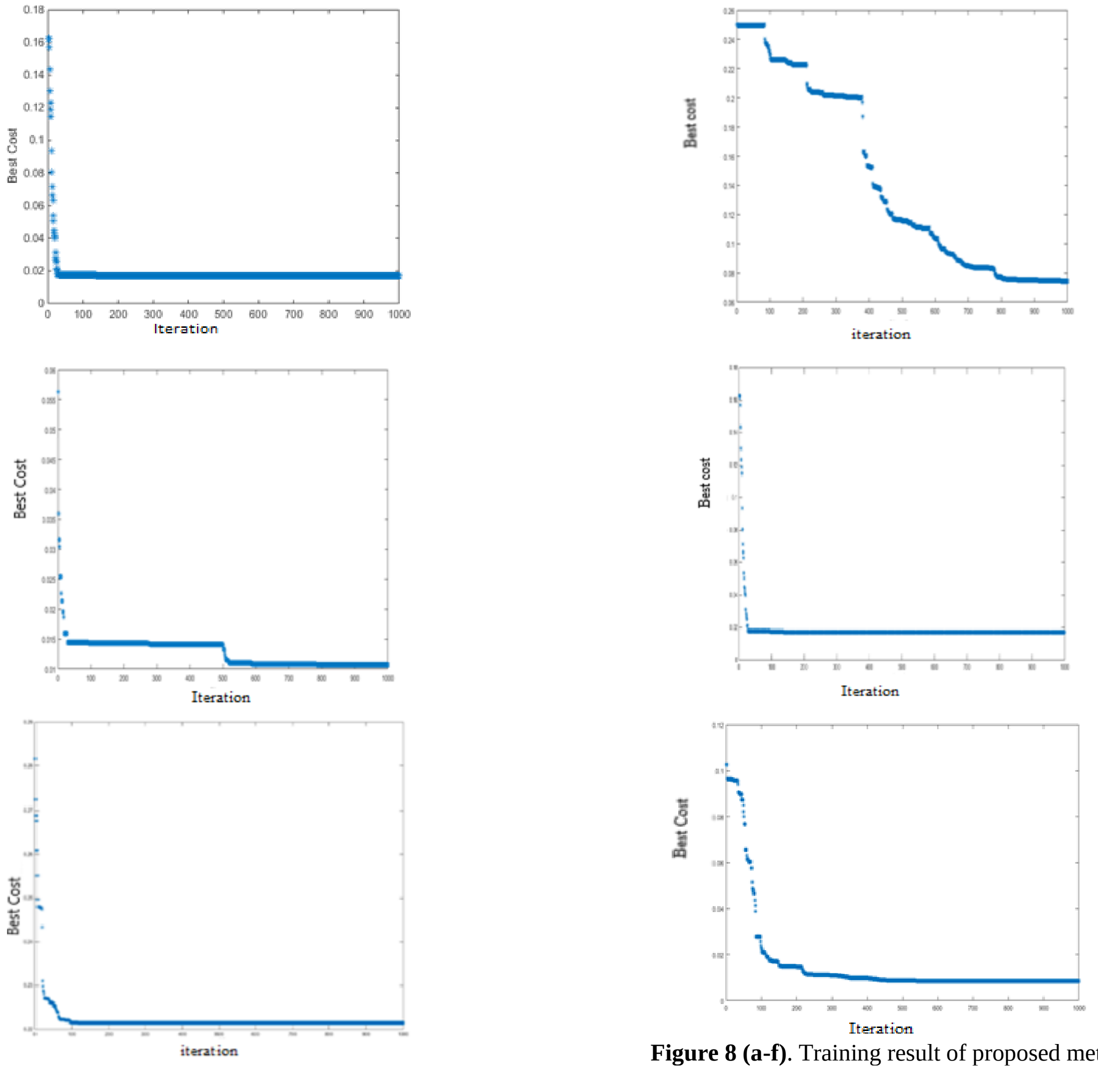

Figure 8 (a-f). Training result of proposed method label and confusion matrix (Equations (13)-(14)). Testing data includes unknown network packets which are classified as legitimate packets or DDoS attacks using trained HMGOGA, elected coefficients and considering cost function. At last, conventional GOA is implemented in similar conditions to make a direct comparison with the proposed model (Table 4).

$$Accuracy\,(train\,/\,test) = \frac{\sum_{i=1}^{N} N_{(predict_i = desire_i)}}{N_T} \times 100 \quad (13)$$

		Actual	
		Positive	**Negative**
Predicted	**Positive**	True Positive (TP)	False Positive (FP)
	Negative	False Negative(FN)	True Negative(TN)

Rate

$$TPR = \frac{TP}{TP+FN} \qquad FPR = \frac{FP}{FP+TN}$$
$$FNR = \frac{FN}{FN+TP} \qquad TNR = \frac{TN}{TN+FP} \quad (14)$$

According to detection sensitivity of DDoS attacks, the confusion matrix is used to prove the stability of the proposed method. The most important metrics in attack detection are TP and FN where TP is the number of attacks correctly classified as attacks and FN is the number of attacks incorrectly classified as normal records. Furthermore, TN and FP are the number of normal records correctly classified as normal records and number of normal records incorrectly classified as attacks (Equation (14)).

As shown in the results, HMGOGA with nonlinear regression cost function has converged efficiently and obtained high-performance accuracy with low FN. Therefore, the proposed model using non-linear cost function has a better performance in comparison with other cost functions. Additionally, HMGOGA outperforms conventional GOA algorithm in every aspect (Table 4). In order to depict the details of HMGOGA algorithm, Mean Cost, Best Cost and Worst Cost of all implemented models are obtained but due to space restriction in this paper, we have just illustrated one of them in Figure 9 to exhibit the different trends of the Worst, Mean and Best populations. According to this figure, the charts are not coincident with one another but all of the 3 charts have a downward trend after a specific iteration, for the generation of each population is based on the prior population.

As shown in Table 4, the proposed nonlinear regression fitted to the model better than other cost functions. The coefficients of cost function are optimized using the implemented meta-heuristic algorithms. The results suggested that HMGOGA based on proposed nonlinear regression has a more accurate and robust performance compared with conventional GOA in case of DDoS detection. The robustness of the proposed model is proved by obtaining low FP and FN.

The receiver operating characteristics (ROC) curve is one of the most important metrics for evaluating the model's performance and it can compare sensitivity versus specificity across a range of values for the ability to predict dichotomous outputs. The area under the ROC curve is another measure of test performance that is shown in Figure 10. The area under curve (AUC) in HMGOGA shows better performance compared to conventional GOA. Indeed, The AUC of HMGOGA depicts the high accuracy and high recall of the proposed model in different thresholds. In order to prove the robustness of the model, some other statistical

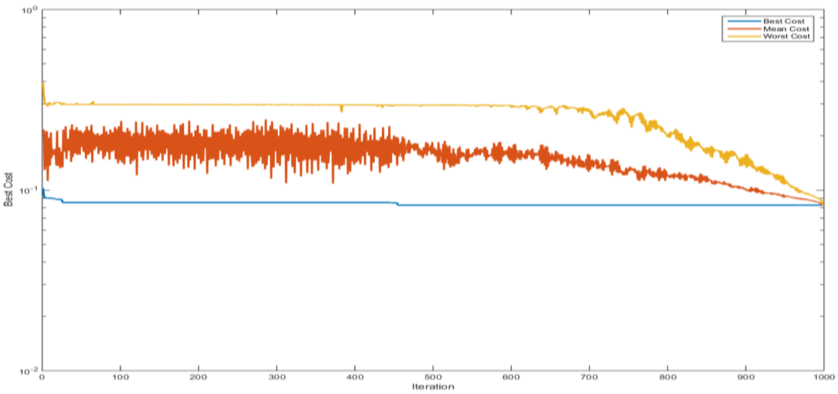

Figure 9. Best, Mean and worst cost of GOA for Ackley function

TABLE 2. Performance analysis

	Performance metrics	Sphere function	Ackley function	Non-linear regression
HMGOGA	Train accuracy:	98.930%	92.552%	99.907%
	Test accuracy:	97.375%	84.015%	**99.3496**%
	FN rate:	0.001	0.001	**0.001**
	FP rate:	0.042	0.264	**0.012**
	TP rate: (Sensivity or recall)	0.998	0.998	0.998
	TN rate:	0.957	0.735	0.988
GOA	Train accuracy:	97.459%	93.742%	98.277%
	Test accuracy:	92.773%	91.436%	95.846%
	FN rate:	0.001	0.038	0.001
	FP rate:	0.118	0.102	0.082
	TP rate: (Sensivity or recall)	0.998	0.961	0.998
	TN rate:	0.881	0.897	0.917

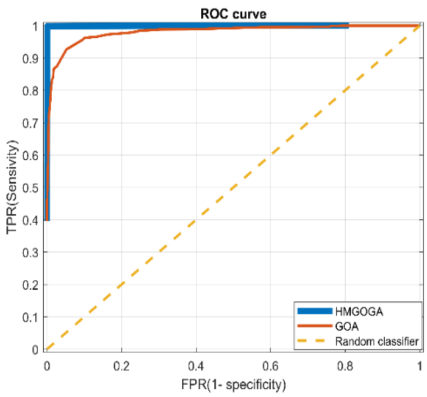

Figure 10. Roc curve for HMGOGA and GOA

test like confidence intervals are calculated in this paper. The robustness of the model is illustrated in Figure 9; where the best cost and mean cost are approximately converged to a single point [41]. According to Equation (15), the obtained accuracy is 99.35% 0.001 by 99% confidence interval (z=2.576).

$$ConfidenceInterval = z \times \sqrt{\frac{error \times (1-error)}{N}} \tag{15}$$

Considering the related research projects on DDoS and intrusion detection systems, many researchers have employed machine learning and meta-heuristic techniques. Hence, some related studies are investigated to compare the performance of our proposed model and validate the efficiency. As shown in Table 5, our proposed method has utilized a novel dataset and meta-heuristic algorithm in DDoS detection scope and achieved a better detection accuracy in comparison with other related research.

TABLE 5. Comparison analysis

References	Dataset	Detection method	accuracy
Bista et al. [42]	CAIDA UCSD DRAPA 2000	Heuristic clustering algorithm and Nave-Bayesian classifier	99.45% , 86.73%
Arivudainambi et al. [22]	NSL-KDD cup	Lion optimization algorithm + Convolutional neural network	98.2%
Sreeram et al.[24]	CAIDA	bio-inspired bat algorithm	94.8%
Proposed method	**CICIDS**	**HMGOGA + Random Forest**	**99.3496%**

5. CONCLUSION

In this paper, a DDoS detection framework has been devised based on the latest meta-heuristic algorithm called GOA in conjunction with a new benchmark dataset called CICIDS and a potent feature selection method called Random Forest. Initially, the most relevant features are extracted from CICIDS dataset using RF feature selection method. The aforementioned dataset consists of 77 features about 75% of which are irrelevant features and are removed from the dataset. Selected features are utilized by GOA algorithm with different cost functions. Considering some weaknesses of GOA: low convergence speed and getting stuck in local minimum, this algorithm is modified and then combined with genetic algorithm (named HMGOGA). As it can be inferred from the results, HMGOGA algorithm confirms better performance in terms of accuracy and robustness. Regarding the novelty of the utilized dataset and meta-heuristic algorithm, the main contributions of this proposed framework is listed below.

1. The Random Forest (RF) feature selection method is applied to our utilized dataset and the 20 most important features among 77 are selected. The performance of the preferred feature subset is validated using DT classifier measures: Mean Absolute Error (MAE) and Correlation Coefficient. High correlation and low MAE have been obtained from our selected features.

2. Low convergence speed and getting stuck in local optimum are two drawbacks of GOA algorithm. In order to overcome these shortcomings, the new SI method is proposed to solve the convergence problem (MGOA). Then, Genetic algorithm is employed to adjust the exploration and exploitation phase and improve the search capability of GOA. The proposed algorithm is called HMGOGA.

3. Two meta-heuristic algorithms (HMGOGA and conventional GOA) are implemented to detect DDoS attacks. HMGOGA and GOA are implemented in similar conditions. The results indicate that the HMGOGA outperforms GOA in terms of detection accuracy and robustness.

4. In order to evaluate the performance and extendibility of the HMGOGA, the proposed framework is implemented using 3 benchmark functions: Sphere, Ackley function and the combined NR-Sigmoid function. The results reveal that NR-Sigmoid function proves to perform better in both HMGOGA and GOA by 99.34 and 95.84 percent test accuracy. In addition, the accuracy of HMGOGA is higher than GOA in all targeting functions. Indeed, nonlinear regression discovered the hidden relation of data more properly.

6. REFERENCES

1. A. Saied, R. E. Overill, and T. J. N. Radzik, "Detection of known and unknown DDoS attacks using Artificial Neural Networks,", ***Neurocomputing*** Vol. 172, (2016), 385-393. https://doi.org/10.1016/j.neucom.2015.04.101

2. O. Osanaiye, K.-K. R. Choo, M. J. J. o. N. Dlodlo, and C. Applications, "Distributed denial of service (DDoS) resilience in cloud: Review and conceptual cloud DDoS mitigation framework," ***Journal of Network and Computer Applications*** Vol. 67, (2016), 147-165. https://doi.org/10.1016/j.jnca.2016.01.001

3. C. Kolias, G. Kambourakis, A. Stavrou, and J. J. C. Voas, "DDoS in the IoT: Mirai and other botnets," ***Computer***, Vol. 50, No. 7, (2017), 80-84. DOI: 10.1109/MC.2017.201

4. J. Wang, M. Yin, and J. J. T. C. S. Wu, "Two approximate algorithms for model counting," ***Theoretical Computer Science*** Vol. 657, (2017), 28-37. https://doi.org/10.1016/j.tcs.2016.04.047

5. C. Wang, T. T. Miu, X. Luo, J. J. I. T. o. I. F. Wang, and Security, "SkyShield: a sketch-based defense system against application layer DDoS attacks," ***IEEE Transactions on Information Forensics and Security,*** Vol. 13, No. 3, (2017), 559-573. DOI: 10.1109/TIFS.2017.2758754

6. Gharvirian, Fateme, and Ali Bohloli. "Neural network based protection of software defined network controller against distributed denial of service attacks." ***International Journal of Engineering***, ***Transactions B: Applications*** Vol. 30, No. 11 (2017), 1714-1722. DOI: 10.5829/ije.2017.30.11b.12

7. A. R. a. Yusof, N. I. Udzir, and A. J. I. J. o. D. E. T. Selamat, "Systematic literature review and taxonomy for DDoS attack detection and prediction,", ***International Journal of Digital Enterprise Technology,*** Vol. 1, No. 3, (2019), 292-315. https://doi.org/10.1504/IJDET.2019.097849

8. Ö. Cepheli, S. Büyükçorak, G. J. J. o. E. Karabulut Kurt, and C. Engineering, "Hybrid intrusion detection system for ddos attacks," ***Journal of Electrical and Computer Engineering*** Vol. 2016, (2016). https://doi.org/10.1155/2016/1075648

9. M. H. Bhuyan, D. Bhattacharyya, J. K. J. S. Kalita, and C. Networks, "E-LDAT: a lightweight system for DDoS flooding attack detection and IP traceback using extended entropy metric," *Security and Communication Networks* Vol. 9, No. 16,(2016) 3251-3270. https://doi.org/10.1002/sec.1530

10. Shamshirband, Shahab, Mahdis Fathi, Anthony T. Chronopoulos, Antonio Montieri, Fabio Palumbo, and Antonio Pescapè. "Computational intelligence intrusion detection techniques in mobile cloud computing environments: Review, taxonomy, and open research issues." ***Journal of Information Security and Applications*** Vol. 55, (2020), 102582. https://doi.org/10.1016/j.jisa.2020.102582

11. Beiki, H., S. M. Seyedhosseini, V. R. Ghezavati, and S. M. Seyedaliakbar. "Multi-objective Optimization of Multi-vehicle Relief Logistics Considering Satisfaction Levels under Uncertainty." ***International Journal of Engineering, Transactions B: Applications,*** Vol. 33, No. 5, (2020), 814-824. DOI: 10.5829/ije.2020.33.05b.13

12 . Fathollahi-Fard, Amir Mohammad, Mostafa Hajiaghaei-Keshteli, and Reza Tavakkoli-Moghaddam. "Red deer algorithm (RDA): a new nature-inspired meta-heuristic." ***Soft Computing*** (2020), 1-29. https://doi.org/10.1007/s00500-020-04812-z

13. Hajiaghaei-Keshteli, Mostafa, Ahmad J Afshari, and Elahe Nasiri. "Addressing the freight consolidation and containerization problem by recent and hybridized meta-heuristic algorithms." ***International Journal of Engineering***, ***Transactions C: Aspects,*** Vol. 30, No. 3, (2017), 403-410.
DOI: 10.5829/idosi.ije.2017.30.03c.10

14. Fathollahi-Fard, Amir Mohammad, Mostafa Hajiaghaei-Keshteli, and Seyedali Mirjalili. "A set of efficient heuristics for a home healthcare problem." ***Neural Computing and Applications,*** Vol. 32, No. 10, (2020), 6185-6205. https://doi.org/10.1007/s00521-019-04126-8

15. I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "A Detailed Analysis of the CICIDS2017 Data Set," in *International Conference on Information Systems Security and Privacy*, 2018, 172-188: Springer. https://doi.org/10.1007/978-3-030-25109-3_9

16. Y. Gu, K. Li, Z. Guo, and Y. J. I. A. Wang, "Semi-Supervised K-Means DDoS Detection Method Using Hybrid Feature Selection Algorithm," ***IEEE Access*** Vol. 7, (2019), 64351-64365. DOI: 10.1109/ACCESS.2019.2917532

17. Ghasemi, J., and J. Esmaily. "A novel intrusion detection systems based on genetic algorithms-suggested features by the means of different permutations of labels' orders." ***International Journal of Engineering, Transactions A: Basics,*** Vol. 30, No. 10, (2017), 1494-1502. DOI: 10.5829/ije.2017.30.10a.10

18. S. M. T. Nezhad, M. Nazari, and E. A. J. I. C. L. Gharavol, "A novel DoS and DDoS attacks detection algorithm using ARIMA time series model and chaotic system in computer networks,"

IEEE Communications Letters, Vol. 20, No. 4,(2016), 700-703. DOI: 10.1109/LCOMM.2016.2517622

19. X. Chen, S. J. I. T. o. I. Yu, and Systems, "A collaborative intrusion detection system against DDoS for SDN," ***IEICE Transactions on Information and Systems,*** Vol. 99, No. 9, 2395-2399, 2016. DOI: 10.1587/transinf.2016EDL8016
20. Singh, Maninder Pal, and Abhinav Bhandari. "New-flow based DDoS attacks in SDN: Taxonomy, rationales, and research challenges." ***Computer Communications,*** (2020). https://doi.org/10.1016/j.comcom.2020.02.085
21. Q. Niyaz, W. Sun, and A. Y. J. a. p. a. Javaid, "A deep learning based DDoS detection system in software-defined networking (SDN)," arXiv preprint arXiv: *1611.07400* 2016. DOI: 10.4108/eai.28-12-2017.153515
22. D. Arivudainambi, V. K. KA, S. S. J. N. C. Chakkaravarthy, and Applications, "LION IDS: A meta-heuristics approach to detect DDoS attacks against Software-Defined Networks," ***Neural Computing and Applications*** Vol. 31, No. 5, (2019), 1491-1501. https://doi.org/10.1007/s00521-018-3383-7
23. M. Yazdani, F. J. J. o. c. d. Jolai, and engineering, "Lion optimization algorithm (LOA): a nature-inspired metaheuristic algorithm," ***Journal of Computational Design and Engineering***, Vol. 3, No. 1, (2016), 24-36. https://doi.org/10.1016/j.jcde.2015.06.003
24. I. Sreeram, V. P. K. J. A. c. Vuppala, and informatics, "HTTP flood attack detection in application layer using machine learning metrics and bio inspired bat algorithm," ***Applied Computing and Informatics***, Vol. 15, No. 1 (2019): 59-66.https://doi.org/10.1016/j.aci.2017.10.003
25. N. Dayal, P. Maity, S. Srivastava, R. J. S. Khondoker, and C. Networks, "Research trends in security and DDoS in SDN," ***Security and Communication Networks,*** Vol. 9, No. 18, (2016), 6386-6411. https://doi.org/10.1002/sec.1759
26. Bala, Ritu, and Ritu Nagpal. "A REVIEW ON KDD CUP99 AND NSL-KDD DATASET." International ***Journal of Advanced Research in Computer Science,*** Vol. 10, No. 2 (2019), 64.

https://doi.org/10.26483/ijarcs.v10i2.6395

27. I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization," in ***ICISSP***, (2018), 108-116. DOI: 10.5220/0006639801080116
28. S. Alelyani, J. Tang, and H. Liu, "Feature selection for clustering: A review," in *Data Clustering*: Chapman and Hall/CRC,(2018),29-60. https://doi.org/10.1007/springerreference_63708
29. A. B. Brahim, M.. Limam, , "Ensemble feature selection for high dimensional data: a new method and a comparative study," ***Advances in Data Analysis and Classification,*** Vol. 12, No. 4, (2018), 937-952. https://doi.org/10.1007/s11634-017-0285-y
30. T. J. Bihl, K. W. Bauer, M. A. J. I. T. o. I. F. Temple, "Feature selection for RF fingerprinting with multiple discriminant analysis and using ZigBee device emissions," ***IEEE Transactions on Information Forensics and Security,*** Vol. 11, No. 8, (2016), 1862-1874. DOI: 10.1109/TIFS.2016.2561902.
31. S. Saremi, S. Mirjalili, and A. J. A. i. E. S. Lewis, "Grasshopper optimisation algorithm: theory and application," ***Advances in Engineering Software***, Vol. 105, (2017), 30-47. https://doi.org/10.1016/j.advengsoft.2017.01.004
32. S. Saremi, S. Mirjalili, S. Mirjalili, and J. S. Dong, "Grasshopper Optimization Algorithm: Theory, Literature Review, and Application in Hand Posture Estimation," in *Nature-Inspired Optimizers*: Springer, (2020), 107-122. https://doi.org/10.1007/978-3-030-12127-3_7
33. Abualigah, Laith, and Ali Diabat. "A comprehensive survey of the Grasshopper optimization algorithm: results, variants, and applications." ***Neural Computing and Applications,*** (2020), 1-24. https://doi.org/10.1007/s00521-020-04789-8
34. Bansal, Priti, Sachin Kumar, Sagar Pasrija, and Sachin Singh. "A hybrid grasshopper and new cat swarm optimization algorithm for feature selection and optimization of multi-layer perceptron." ***Soft Computing,*** (2020), 1-27. https://doi.org/10.1007/s00500-020-04877-w
35. J. Luo, H. Chen, Y. Xu, H. Huang, and X. J. A. M. M. Zhao, "An improved grasshopper optimization algorithm with application to financial stress prediction," ***Applied Mathematical Modelling***, Vol. 64, (2018), 654-668. https://doi.org/10.1016/j.apm.2018.07.044
36. Joshi, Susheel Kumar, and Jagdish Chand Bansal. "Parameter tuning for meta-heuristics." ***Knowledge-Based Systems*** 189 (2020), 105094. https://doi.org/10.1016/j.knosys.2019.105094
37. H. Binol, I. Guvenc, E. Bulut, and K. J. E. L. Akkaya, "Hybrid evolutionary search method for complex function optimisation problems," ***Electronics Letters,*** Vol. 54, No. 24, (2018), 1377-1379. DOI: 10.1049/el.2018.6506
38. D. Ackley, *A connectionist machine for genetic hillclimbing*. Springer Science & Business Media, (2012). https://doi.org/10.1007/978-1-4613-1997-9
39. M. Babagoli, M. P. Aghababa, and V. J. S. C. Solouk, "Heuristic nonlinear regression strategy for detecting phishing websites," ***Soft Computing,*** Vol. 23, No. 12, (2019), 4315-4327. https://doi.org/10.1007/s00500-018-3084-2
40. S. Mirjalili and A. J. A. i. e. s. Lewis, "The whale optimization algorithm," ***Advances in engineering software,*** Vol. 95, (2016), 51-67. https://doi.org/10.1016/j.advengsoft.2016.01.008
41. Haddadi, Mohamed, and Rachid Beghdad. "A Confidence Interval Based Filtering Against DDoS Attack in Cloud Environment: A Confidence Interval Against DDoS Attack in the Cloud." ***International Journal of Information Security and Privacy (IJISP)***, Vol. 14, No. 4, (2020), 42-56. DOI: 10.4018/IJISP.2020100103
42. S. Bista and R. J. J. o. I. S. Chitrakar, "DDoS Attack Detection Using Heuristics Clustering Algorithm and Naïve Bayes Classification," ***Journal of Information Security,*** Vol. 9, No. 01, (2017), 33. DOI: 10.4236/jis.2018.91004

Persian Abstract

چکیده

امنیت سایبری به دلیل گسترش تهدیدات سایبری و آزار و اذیت‌های اینترنتی به محیطی وحشیانه و شرورانه تبدیل شده‌است. حمله انکار سرویس توزیع شده (DDoS) تهدیدی در شبکه است که منابع قربانیان را به خطر می اندازد. با توجه به نقش قابل توجه الگوریتم های بهینه سازی در شناسایی بسیار دقیق، قابلیت انطباق با حملات شبکه و نرخ هشدار کاذب قابل قبول، مطالعه حاضر یک روش ترکیبی مبتنی بر الگوریتم بهینه سازی ملخ اصلاح شده و الگوریتم ژنتیک (HMGOGA) برای شناسایی و جلوگیری از حملات DDoS پیشنهاد کرده است. HMGOGA بر معایب الگوریتم GOA سنتی از جمله سرعت همگرایی کم و گیر افتادن در بهینه محلی غلبه می کند. در این مقاله ، از الگوریتم پیشنهادی برای شناسایی حملات DDoS از طریق شبیه سازی مدل رگرسیون غیرخطی (NR) استفاده شده است. به منظور دستیابی به این منظور ، در ابتدا مهمترین ویژگیهای ترافیک شبکه با استفاده از روش جنگل تصادفی (RF) استخراج می شود. با حذف 55 ویژگی بی ربط از مجموع 77 ویژگی، ویژگی‌های منتخب نقش اساسی در عملکرد مدل پیشنهادی ایفا کرده اند. برای صحت سنجی کارایی ، همبستگی زیاد ویژگی های انتخاب شده با درخت تصمیم (DT) اندازه گیری شده است. متعاقباً ، HMGOGA با دو تابع هزینه معتبر و یک تابع هزینه پیشنهادی دیگر آموزش داده می شود که به آن امکان می دهد ترافیک مخرب را به درستی تشخیص دهد. جهت اعتبار سنجی و قابل استفاده بودن مدل پیشنهادی با دو تابع هزینه معتبر (عملکرد Sphere و Ackley) مقایسه می شود. نتایج تجربی ثابت کرده است که HMGOGA مبتنی بر NR-sigmoid به ترتیب با 99.90٪ و 99.34٪ دقت آموزش و آزمون، به ترتیب از سایر مدلهای پیاده سازی شده و روشهای GOA معمولی عملکرد بهتری داشته است.