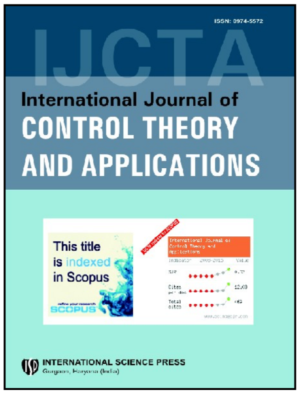

ISSN: 0974-5572

IJCTA

International Journal of

CONTROL THEORY AND APPLICATIONS

This title is indexed in Scopus

INTERNATIONAL SCIENCE PRESS

International Journal of Control Theory and Applications

ISSN : 0974-5572

© International Science Press

Volume 10 • Number 26 • 2017

Genetic Bat for Distributed Denial of Service Attack Classification in Cloud

S.Velliangiri and J. Premalatha

Department of CSE, Sri Venkatesa Perumal College of Engineering and Technology, Puttur, India. Velliangiri.s.2009@gmail.com
Professor, Kongu Engineering College, Erode, India

Abstract: In recent years, cloud computing offers various services for the people and also it has been attacked seriously through intruders; however people may not reveal to be responsible. Intrusion Detection Systems (IDS) is a common aspect in network protection infrastructure and from several attacks it protects systems. In computer linked internet, Distributed Denial of Service (DDoS) attacks may be the major challenge. Here, Radial Basis Function Neural Network (RBF-NN) detector is provided to find DDoS attacks. On the basis of understanding and skill, several training algorithms are decided for RBF-NNs to initiate a programmed network structure. Trial and error method are employed to organize resulting network. Genetic Algorithm (GA) and Bat Algorithm (BA) are employed to organize RBF-NN in an automatic manner. The efficiency of this proposed method is certified through simulation result.

Keywords: Cloud Computing, Intrusion Detection Systems (IDS), Distributed Denial of Service (DDoS), Radial Basis Function (RBF), Neural Network (NN), Genetic Algorithm (GA) and Bat Algorithm (BA).

1. INTRODUCTION

Global, on-demand Internet access with cloud computing allows access to resources and services with minimal interaction with service providers. Cloud computing has characteristics like on-demand service, access to extensive network, resource pooling, and metered services which allows users to focus on their own businesses processes and the cloud service provider manages the computing resources. The cloud model significantly lowers business costs and ensures accessibility and adaptability of computing resources [1]. The cloud provides service in three forms of service: as infrastructure providing computing resources, as platform facilitating developers with providers and as software providing users with applications running on cloud. As cloud gets together various players into play, the users, providers, developers, it is essential to have strong security mechanisms to maintain the confidentiality and integrity of the data in the cloud.

An intrusion is defined as an attempt to compromise the Confidentiality, Integrity, and Accessibility of the security mechanisms of a system. Attackers trigger intrusions to access cloud resources illegally. Intrusion detection helps in identifying the intrusions by monitoring and analysing the traffic in the network. Intrusion Detection Systems (IDS) can also be classified into following categories on the basis of the detection approaches:

Misuse detection (or signature based detection): these systems work by matching user activity with stored signatures of known attacks. Such detection systems use a predefined knowledgebase to check whether the new network connection is in that knowledge database. If yes, the IDS consider this connection as a possible attack and then block it. Anomaly detection (or Behavior detection): In this case, the system learns the characteristics of normal user activities and then uses such characteristics to judge whether new user's activity is normal or not [2].

A Denial of Service (DoS) attack prevents legitimate users from accessing the network resource. A Distributed DoS (DDoS) attack is a coordinated attack through many compromised units which severely restricts the access of services to the users. As the attackers may launch its offensive action through thousands of IP addresses, it becomes difficult to separate legitimate traffic from malicious traffic. The DoS and DDoS attacks not only cause losses but also leads to lower productivity and services and increases downtime [3]. The DoS and DDoS attacks are the most common but fatal type of attack on CSPs which are working hard to prevent, monitor and mitigate these types of attacks as the frequency of these types of attacks have risen sharply in the last few years. DDoS are directed at service provider's infrastructure can be very damaging. In cloud computing, the DoS or DDoS attack is when a machine or network resources unavailable to its intended users. An IDS for predicting the DDoS attacks in cloud will significantly improve the security of the cloud system.

In general, the problem of mapping tasks on apparently unlimited computing resources in cloud computing belongs to a category of problems known as Non-deterministic Polynomial (NP)-hard problems. There are no algorithms which may produce optimal solution within polynomial time for such kind of problems. Solutions based on exhaustive search are not feasible as the operating cost of generating schedules is very high. Meta-heuristic based techniques deal with these problems by providing near optimal solutions within reasonable time. Meta-heuristics have gained huge popularity in the past years due to its efficiency and effectiveness to solve large and complex problems. In this work, the author present a GA, BA and RBF-NN based DDoS attack in cloud computing. Section 2 reviews related work in literature. Section 3 describes methods used and Section 4 discusses experiments results. Section 5 concludes the work.

2. RELATED WORKS

Ravale *et al.,* [4] had discussed on clustering, classification and association rule discovery. Proposed hybrid method was employed to unite K Means clustering algorithm and Radial Basis Function (RBF) kernel function. This method had minimized related attributes with all data points and offered greater performance on detection rate and accurateness as Knowledge Discovery and Data Mining (KDDCUP'99) data set had been applied.

Benaicha *et al.,* [5] acquired GA approach to improve initial population detection and to detect several network intrusions in an effective manner. Network Security Laboratory-Knowledge Discovery and Data Mining (NSL-KDD99) benchmark dataset were employed to detect abuse activities in testing phase. Detection rate performance had been improved through combining IDS with GA and false positive rate was minimized.

Mizukoshi & Munetomo [6] introduced an available, real-time traffic pattern analysis to detect and reduce DDoS attacks on GA technique on an infrastructure of Hadoop distributed processing. Efficiency was revealed for a scalable DDoS protection system via experiments.

Enache & Sgârciu [7] introduced an IDS model and it composed of two stages like information gain based feature selection and SVM detection. SVM had drawbacks that the consequences were persuaded through input parameters and hence Swarm Intelligence, BA merits were utilized to attain an enhanced classifier and randomization with Leìvy flights. NSL-KDD dataset was exploited for testing and consequence showed that proposed technique outperformed BA, Artificial Bee Colony (ABC) and Particle Swarm Optimization (PSO).

Enache & Sgârciu [8] provided a network anomaly IDS and it combined SVMs classifier with enhanced BA. SI algorithm's binary version constructed a wrapper based feature selection and basic version to select

SVM input parameters. NSL-KDD dataset was employed for testing and revealed that this model had performed well over SVM or techniques based on PSO and BA for attack detection and false alarm rate.

3. METHODOLOGY

Three-layer feed-forward network is the RBF-NN composed as a single hidden layer and it is a nonstop process with accurate arbitrary. RBF-NN features are structure-adaptive determination and independent initial output value. In earlier work, nonlinear learning algorithms was employed however in existing work, linear learning algorithms had been exploited with characteristics like optimum approximation, global optimum, etc., and additionally greater non-linear accuracy was maintained. For conventional classification issue the RBF-NN is exploited. Optimization of RBF-NN with GA is proposed here.

(A) Radial Basis Function (RBF) Network

Multiple RBF network layers perform several tasks and it divides hidden and output layer optimization. This network is intended on center determination of RBFs such as fixed, and centers selection through several learning schemes.

The adjustable parameters within a RBF network that effect the classification accuracy depends upon the factors like basic functions used, location of the centre of the basis function, width of the basis function, and the weights connecting the hidden RBF units to the linear output units. The RBF network when fed with the input comprising of the traffic data, Each Hidden neuron symbolizes the traffic as an attack or normal and output combines to classify it as normal or equivalent attack [9].

(B) Genetic Algorithm (GA)

GA impressionist’s biological evolution as an issue resolving scheme and it tracks Darwinian’s principle of evolution and survival fittest to enhance candidate’s population solution to a fixed fitness. Progression and natural selection of chromosome-like data structure is employed by GA and progress the chromosomes through operators such as selection, recombination and mutation. Chromosomes population is produced randomly and it initiates the GA systems representing solution of each problem are deemed as a candidate solution. Bits, characters or numbers are set for several locations from all chromosome and these positions are named as genes. On the basis of favoured solution, goodness of all chromosomes is computed through assessment value and it is termed as Fitness Function. Cross over simulates natural reproduction and mutation of species is carried out Mutation. Chromosome selection is predisposed to fittest chromosomes for survival and combination [10].

(C) Bat Algorithm (BA)

Echolocation capability of bat engrossed researcher’s attention from multiple fields. Loud and small pulse of sound is yielded by bats through hits an object and echo return back after fraction of time to bats ear so that it compute the distance between object and bat. Throughout iteration sequence, position, velocity, and frequency vector is comprised for an artificial bat. In the BA, the artificial bats can or updated position vectors) within the continuous real domain [11].

(D) Proposed GA-BA-RBF algorithm

The main advantage of using GA for optimizing RBF network are that the RBF parameters can be easily encoded in the chromosome, defining fitness function, and the use of genetic operators to achieve optimal solution. The proposed GA-RBF optimization algorithm automatically adjusts network structure and connection weights.

The GA-RBF neural network algorithm basis step is descried as follows [12].

Step 1: Set the RBF neural network, according to the maximum number of neurons in the hidden layers; use K-clustering algorithm to obtain the center of basis function use formula $\sigma = \frac{d}{\sqrt{2s}}$ to calculate the width of the center.

Step 2: Set the parameters of the GA, the population size, the crossover rate, mutation rate, selection mechanism, crossover operator and mutation operator, the objective function error, and the maximum number of iterations.

Step 3: Initialize populations *P* randomly; its size is *N* (the number of RBF neural network is *N*); the corresponding network to each individual is encoded by formula $\begin{matrix} c_1c_2...c_s w_{11}w_{21}...w_{s1}w_{12}w_{22} \\ ...w_{s1}...w_{1m}w_{1m}...w_{sm}\theta_1\theta_2...\theta_m \end{matrix}$.

Step 4: Use the training sample to train the initial constructed RBF neural network, whose amount is *N*; use formula $e = \sum_{k=1}^{n}\left(t_k - y_k\right)^2$ to calculate the network's output error *E*.

Step 5: According to the training error *E* and the number of hidden layer neurons *s*, use formula $F = C - E \times \frac{s}{s_{\max}}$ to calculate the corresponding chromosome fitness to each network.

Step 6: According the fitness value, sort the chromosome; select the best fitness of the population, denoted by F_b; verify $E<E_{\min}$ or $G \geq G_{\max}$; if yes, turn to Step 9; otherwise turn to Step 7.

Step 7: Select several best individuals to be reserved to the next generation New *P* directly.

Step 8: Select a pair of chromosomes for single-point crossover, to generate two new individuals as members of next generation; repeat this procedure, until the new generation reaches the maximum size of population *Ps*; at this time, the coding will be done separately.

Step 9: Mutate the population of new generation; binary coding part and real number coding part should use different mutation strategies. Then the new population is generated; set *P* = New *P*, *G* = *G* + 1; return to Step 4.

Step 10: Get the optimal neural network structure, and the iteration of genetic algorithm is terminated, which means the optimizing stopped.

Step 11: The new neural network's weight learning is not sufficient, so use Least Mean Square (LMS) method to further learn the weights. End of the algorithm.

The significance of establishing new model is that to optimize neural network structure, to determine the number of hidden layer neurons and the center of the basis function, to optimize the connection weight and threshold, in order to improve the training speed and convergence, to save network running time, and then to improve the operating efficiency of network and the ability of dealing with problems.

The process of training RBF network using BA [13] chooses the optimal parameters of the weights between the hidden and output layer (w) and the parameter spread (α), Centers hidden layer (μ) and basis of the cells in a layer output (β). Determination of number of cells in the hidden layer is vital in RBF network as it affects the speed of convergence. If the number of cells is low then the convergence speed is slow and if large number of cells, then complexity of the network structure is high. In this work, the fitness function is based on the Mean Squared Error (MSE). The BA every bat evaluated based on a scale MSE as well as dependence on the values of w and μ and ß and α.

Step 1: BA is initializes and passes the best weights to RBF

Step 2: Load the training data

Step 3: While MSE < Stopping Criteria

Step 4: Initialize all BA Population

Step 5: BA Population finds the best weight and ε

$$v_i^t = v_i^{t-1} + f_i\left(x_i^{t-1} - x_{cgbest}\right)$$

$y = f(x) = \sum_{i=1}^{k} w\phi_i(x) + \beta_i$ and

$$\phi_i(x) = \exp\left[\frac{\left\|x-\mu_i\right\|^2}{2\sigma_i^2}\right].$$

Step 5: RBF is run using the weights and ε obtained by BA

Step 6: BA iterates computing the best possible weight, α, ß, μ till convergence.

Step 7: End While.

In figure 1 shows the steps of GA and BA for training RBF network. The traffic data is given as input, and the algorithm iterates and creates the parameters for each of the RBF network.

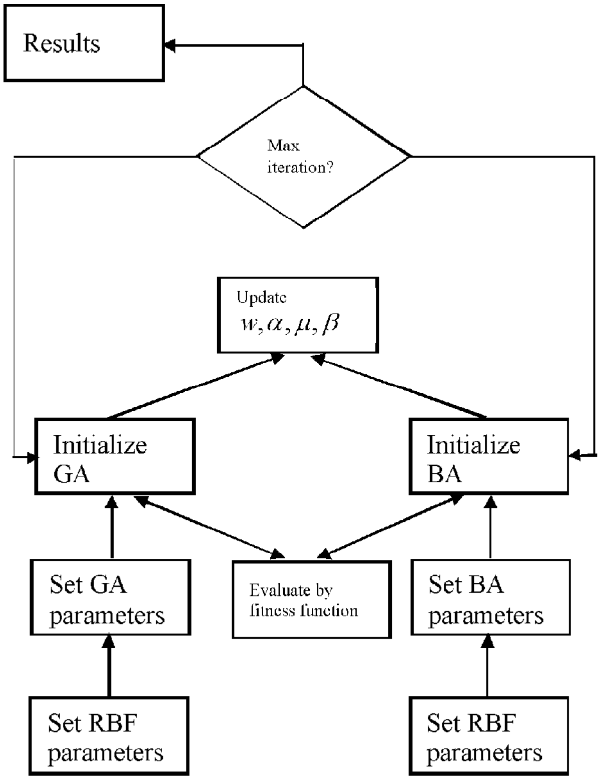

Figure 1: Comparison between BA and GA for training RBF networks

4. RESULTS AND DISCUSSION

In simulation experiments, RBF and GA-BAT-RBF methods were assessed. The classification and false positive rate are shown in table 1 & 2 and figure 2 & 3.

Table 1
Classification Accuracy

Number of second	*RBF*	*GA-BAT-RBF*
2	98.13	98.78
4	98.47	98.92
6	98.73	99.04
8	98.84	99.23
10	98.83	99.42

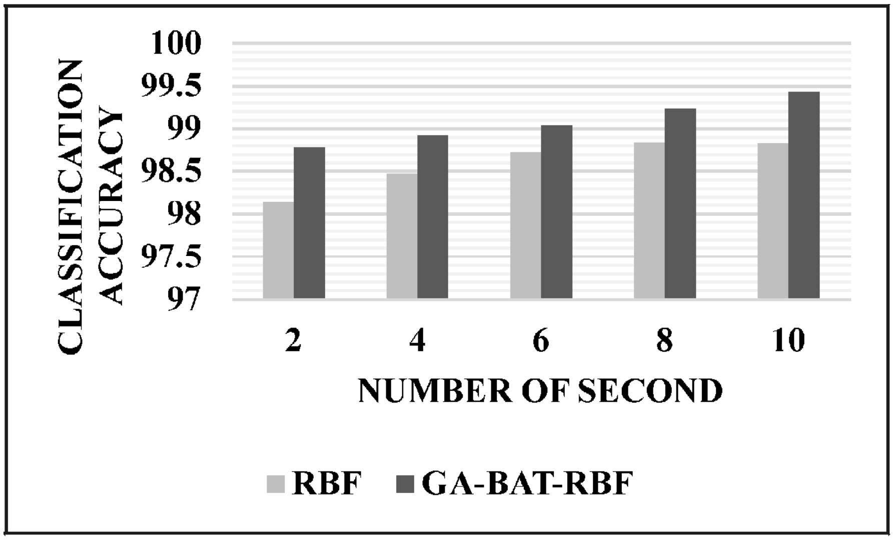

Figure 2: Classification Accuracy

From the table 1 and figure 2, it can be noticed that the GA-BAT-RBF has greater classification accuracy by 0.66% for 2 numbers of second, by 0.45% for 4 numbers of second, by 0.31% for 6 numbers of second, by 0.39% for 8 numbers of second and by 0.59% for 10 numbers of second when compared over RBF.

Table 2
False Positive Rate

Number of second	*RBF*	*GA-BAT-RBF*
2	0.1789	0.1623
4	0.1727	0.1604
6	0.1683	0.1582
8	0.1581	0.1532
10	0.1576	0.1502

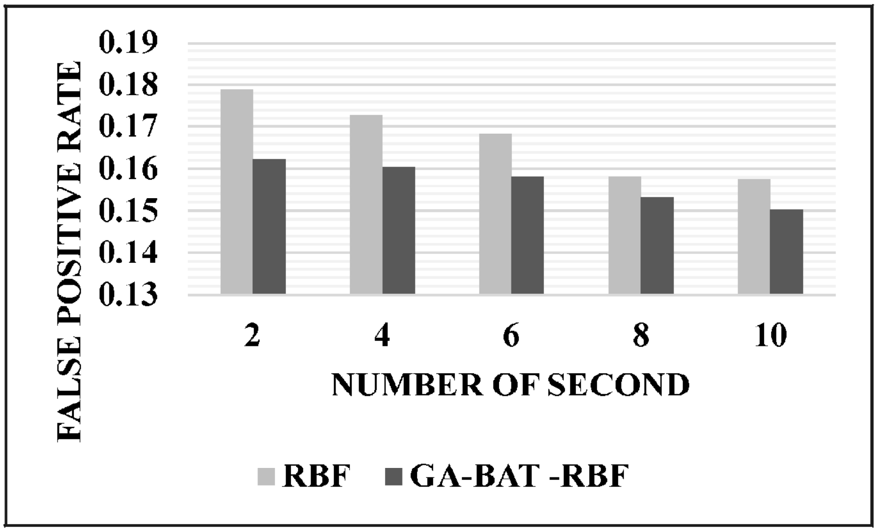

Figure 3: False Positive Rate

From the figure 3, it can be noticed that the GA-BAT-RBF has lower false positive rate by 9.73% for 2 number of second, by 7.38% for 4 number of second, by 6.18% for 6 number of second, by 3.14% for 8 number of second and by 4.8% for 10 number of second when compared over RBF.

5. CONCLUSION

All IT organization's choice is cloud computing since it offers the service as flexible and pay-peruse to its users. However, this environment is susceptible to intruders because of its open and allocated architecture and thus its success is not simple. Generally basic system is IDS to identify attacks on cloud. Based on GA and BA (GA-BA-RBF), an optimized RBF-NN algorithm is presented here. GA optimizes the weights and structure of RBF-NN. Weight, center, and width values of RBF are optimized through BA system. It is instigated by defining its parameter. In BAT destination, RBF weights are encoded and calculate the fitness of all the solutions. Consequences showed that the GA-BAT-RBF has greater classification accuracy by 0.66% for 2 number of second, by 0.45% for 4 number of second, by 0.31% for 6 number of second, by 0.39% for 8 number of second and by 0.59% for 10 number of second when compared over RBF.

REFERENCES

[1] Carlin, A., Hammoudeh, M., &Aldabbas, O. (2015). Defence for Distributed Denial of Service Attacks in Cloud Computing. Procedia Computer Science, 73, 490-497.

[2] Chauhan, K., & Prasad, V. (2015). Distributed Denial of Service (DDoS) Attack Techniques and Prevention on Cloud Environment. International Journal of Innovations & Advancement in Computer Science, 210-215.

[3] Kalra, M., & Singh, S. (2015). A review of metaheuristic scheduling techniques in cloud computing. Egyptian Informatics Journal, 16(3), 275-295.

[4] Ravale, U., Marathe, N., &Padiya, P. (2015). Feature selection based hybrid anomaly intrusion detection system using K means and RBF kernel function. Procedia Computer Science, 45, 428-435.

[5] Benaicha, S. E., Saoudi, L., Guermeche, S. E. B., &Lounis, O. (2014, August). Intrusion detection system using genetic algorithm. In Science and Information Conference (SAI), 2014 (pp. 564-568). IEEE.

[6] Mizukoshi, M., &Munetomo, M. (2015, May). Distributed denial of services attack protection system with genetic algorithms on Hadoop cluster computing framework. In Evolutionary Computation (CEC), 2015 IEEE Congress on (pp. 1575-1580). IEEE.

[7] Enache, A. C., &Sgârciu, V. (2015, May). Anomaly intrusions detection based on support vector machines with an improved bat algorithm. In Control Systems and Computer Science (CSCS), 2015 20th International Conference on (pp. 317-321). IEEE.

[8] Enache, A. C., & Sgârciu, V. (2015). An improved bat algorithm driven by support vector machines for intrusion detection. In International Joint Conference (pp. 41-51). Springer International Publishing.

[9] Hoque, M. S., Mukit, M., Bikas, M., &Naser, A. (2012). An implementation of intrusion detection system using genetic algorithm. arXiv preprint arXiv:1204.1336.

[10] Rajalaxmi, R. R., & Ramesh, A. (2014). Binary Bat Approach for Effective Spam Classification in Online Social Networks. Australian Journal of Basic and Applied Sciences, 8(18), 383-388.

[11] Kandeeban, S. S., & Rajesh, R. S. (2012). A Frame of Intrusion Detection Learning System Utilizing Radial Basis Function. International Journal of Modern Education and Computer Science, 4(1), 19.

[12] Jia, W., Zhao, D., Shen, T., Su, C., Hu, C., & Zhao, Y. (2014). A new optimized GA-RBF neural network algorithm. Computational intelligence and neuroscience, 2014, 44.

[13] Talal, R. (2014). Comparative study between the (ba) algorithm and (pso) algorithm to train (rbf) network at data classification. International Journal of Computer Applications, 92(5).