A Hybrid Algorithm Using Ant Colony Optimisation and Cuttle Fish Algorithm for Feature Selection of Intrusion Detection

V.R.Balasaraswathi, Dr.M.Sugumaran

Abstract —Ant Colony Optimisation (ACO) and cuttle Fish Algorithm (CFA) are the familiar algorithms used to solve many optimisation problems. Intrusion Detection System (IDS) is software used to detect intruders entering into a network. IDS contain many steps such as Pre-processing, Feature Selection (FS), Optimization Model and Classification. FS is an important step in IDS. Many algorithms have been used for feature selection. This paper proposes an hybrid algorithm by combining ACO and CFA for FS.ACO performs global search in a better way and hence it is combined with CFA to generate efficient solution. ACO is used to generate initial solution for CFA. With the best initial solution generated by ACO, CFA generates an optimal feature subset.

Keywords— Ant Colony Optimisation, Cuttle Fish Algorithm, Feature Selection, Optimisation Intrusion Detection System,

Introduction

IDS is used in a network to detect the unauthorised person entering into it [1].A different methods of IDS are Network Intrusion Detection System (NIDS) and Host Based Intrusion Detection System (HIDS). NIDS is organized in a network which monitors the traffic entering and leaving the network [2]. HIDS is deployed on all the devices in a network which monitors the host and the network traffic. It has the advantage of detecting the malicious activities which originates in a host itself [3]. IDS is also classified into signature based IDS, anomaly based IDS and hybrid IDS. Signature based monitors the packets enter inside the network and compares with the database of signatures with known attacks [4].Anomaly based detects the malicious activities which occurs inside the host and the network [5]. Hybrid IDS is the combination of signature based and anomaly based IDS.

Many algorithms such as bio-inspired and non bio-inspired algorithms are used for IDS. Recently bio-inspired algorithms such as evolutionary, genetic and swarm based algorithms are used widely. Swarm based algorithms gives better results when compared with others. A hybrid algorithm using Particle Swarm Optimisation (PSO) and Support Vector Machine (SVM) is used for developing IDS [6]. Swarm intelligence algorithms performs well than the other algorithms [7]. Simplified swarm optimisation with weighted local search strategy is used [8]. Searching process and the overall accuracy is improved. K-means algorithm based on PSO is used [9].PSO is the most intelligent algorithm improves the global search. Support vector machine is combined with kernel principal component analysis (KPCA) and genetic algorithm (GA) is used [10] . SVM reduces the dimension and the training time. Self Organised ant based clustering is used which improves the overall performance [11]. KPCA is combined with chaotic PSO and SVM is used to develop IDS [12].The accuracy and computational time is increased and also training time is reduced. The survey of IDS gives the detailed description of IDS and its algorithms [13]

The various swarm intelligence algorithms were used for developing the IDS. The algorithms such as ACO, SSO, PSO, KPCA, K-means clustering, GA, etc were used. The ACO performs the global search well in a large search space.ACO is used in many research areas such as image processing, machine learning, medical diagnosis, etc.

This paper is organized as follows. Section 2 describes the feature selection. Section 3 discusses about the ACO. Section 4 illustrates the CFA. Section 5 gives the ACO-CFA algorithm. Section 6 discusses about experimental results and section 7 gives the conclusion

2. Feature Selection

FS is the main and the most important step in the design of IDS.FS is the process of reducing the dataset, extracting the relevant and efficient features and removing the noisy features in the dataset. FS is classified into filter, wrapper and hybrid methods. Filter method is a traditional method which does not use classifier for training process. It uses the measures such as attribute ratio, distance, correlation coefficient, similarity, etc are used. Wrapper method is complex and most widely used and it uses classifier for training. Hybrid method is the combination of both.

Many swarm intelligence algorithms are used for FS. ACO, SSO, Bee algorithm, Cuttle fish algorithm (CFA), PSO, genetic algorithm, artificial bee colony algorithm, bat algorithm, etc are used for FS. Recently many metaheuristic algorithms are used for selecting the efficient features. Bee algorithm is improvised [14] by using the membrane computing concept for FS. ACO is used for selecting features and support vector machine is used as a classifier [16]. Detection rate is improved by ACO.ACO is used for FS as well as for classification [17]. Artificial bee colony algorithm is used [18].The overall performance is improved. Bat algorithm is used to reduce the computational time [19].ACO and SVM is used to increase the detection rate and accuracy [20]. Weight is assigned to features and then ACO is applied to select the features. CFA is used to select the relevant features [15] and accuracy have been improved.

3. Ant Colony Optimisation

In 1990, Dorigo was introduced the ACO concept [21]. It is the nature inspired heuristic algorithm which is mostly used to solve optimisation problems with good computational time. The shortest path is found by the ant by depositing a chemical substance pheromone. The ants move in the direction where the pheromone content is rich. The shortest path is crumbled over time and the traversal rate will be more in the shortest path. Pheromone model is the main component of ACO algorithm. First the pheromone values are initialised and then solution is constructed. To solve the optimisation problems, two steps are required.

1. Using the pheromone model, candidate solution is constructed.
2. The pheromone values are modified by using candidate solution to get good solutions.

ACO is used in networks to find the shortest path. The other applications of ACO are scheduling problems, assignment problems, sequential ordering problems, vehicle routing problems and travelling salesman problems. The complexity is very less in ACO algorithm, hence it is used to solve complex and NP complete problems.

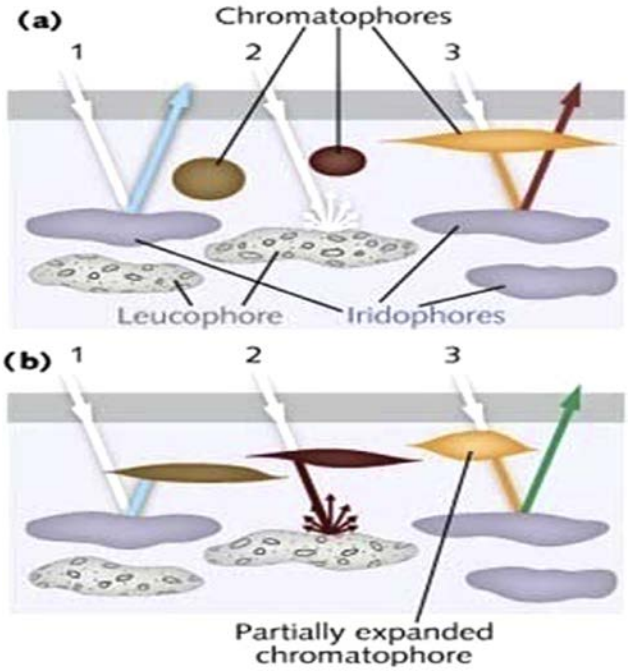
Steps in ACO

1. Initialise the pheromone values.
2. N ants are generated.
3. A subset is constructed for each ant.
4. Solution is constructed.
5. If solution is constructed, pheromone value is updated locally, goto step 3
6. If solution is not constructed, constructed solution is evaluated
7. Pheromone value is updated globally
8. If termination condition is met, the best solution is returned, else goto step2.

The different updates of pheromone value gives the way to obtain optimal solution.

4. Cuttle Fish Algorithm

CFA is a metaheuristic algorithm. It works based on colour changing behaviour . The colours and patterns of cuttlefish are obtained by reflected light of the skin layers. Three layers namely chromatophres, iridophores and leucophores are present in the cuttle fish. Iridophores and leucophores are called as reflecting cells. Light is reflected by chromatophores or by reflecting cells or a combination of both. Skin layers are shown in figure1.



Chromatophores are groups of cells that include an elastic saccule that holds a pigment. When the muscles contract, they stretch the saccule allows the pigment inside to cover a larger surface area. When the muscles relax, the saccule shrinks and hides the pigment. Iridophores are found in the next layer under the chromatophores. Iridophores are responsible for producing the metallic looking greens, blues and golds seen. Works by reflecting light, used to conceal organs and communication. Leucophores are flattened, branched cells scatters and reflects incoming light. Reflects the predominant wavelength of light in the environment. In white light they will be white, while in blue light they will be blue. Provides the ability to blend into its environment.

The appearance of the cuttlefish depends on which skin elements affect the light incident on the skin. CFA mimics the working behaviour of three layers of skin and it is shown by six cases. Six cases are shown in figure 2.

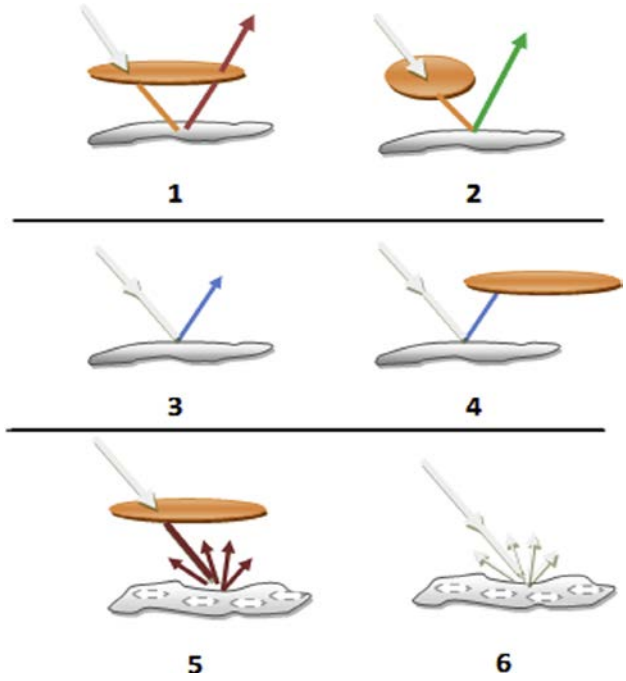


Figure 2 six cases of skin layers

Reflection and visibility are the two main processes. Reflection process is the mechanism which reflects the incoming light and it can by any case of the six cases. Visibility represents the matching pattern that the cuttlefish try to simulate the patterns appear in its environment. CFA divides the population into four groups and each group works independently sharing only the best solution. New solution is obtained by reflection and visibility. CFA starts with random solutions for initializing the population.

Steps of CFA:

1. Initialize the population with random solutions, calculate and keep the best solution and the average value of the best solution.

2. Use interaction operator between chromatophores and iridophores cells in case 1 and 2, to produce a new solution based on the reflection and the visibility of pattern.

3. Use iridophores cells operators in case 3 and 4 to calculate new solutions based on the reflected light coming from best solution and the visibility.

4. Use leucophores cells operator in case 5 to produce new solution by reflecting light from the area around the best solution and visibility of the pattern (local search).

5. ACO-CFA

ACO performs the exploration even in a large search space in a better way. In CFA the initial solution is generated randomly. In the proposed work, the initial solution is generated by ACO for CFA. With the efficient initial solution generated by ACO, CFA obtains the optimal subset of features for IDS.

The size of the ants is set to the number of features and a classifier is used to evaluate the features selected by the ants. Initially a population of ants N is set equal to number of features. Pheromone value is assigned to all features. For the ants to begin the search, each ant is assigned a feature subset consisting of random combination of features. Each ant selects feature based on the probability as specified in eq.1

$$p_i=\tau_i . \Delta\tau_i \quad (1)$$

Where τ_i is the pheromone value of the feature i and $\Delta\tau_i$ is the proportion of ants that have selected this feature.

Whenever ant selects a feature, the pheromone value of the feature is updated using eq.2.

$$\tau_i = (1-\varphi)\ \tau_i + \varphi.\tau_0 \quad (2)$$

Where φ is the parameter of relative importance and takes values from 0 to 1.

After all ants have finished a run, the feature subsets selected by the ants are passed to CFA. Feature subsets leading to ineffective solutions are not selected. The features leading to high-quality solutions are selected. The global pheromone update is done by using the equation 3.

$$\tau_i = (1-\varrho)\ \tau_i + (\varrho\ .1/\ P_{optimal}) \quad (3)$$

Where ϱ is the pheromone evaporation rate. $P_{optimal}$ is the subset which has highest accuracy

Steps of ACO-CFA

1. l=1
2. ACO is initialised

   Number of ants A is assigned to number of features F

   For i =1 to F, pheromone value is assigned to features τ_i
3. CFA is initialised
4. Ant:

Repeat

For each feature i,probability of selecting feature i is calculated using eq 1.

Table 1 KDDCUP'99 Dataset experimental values

S. No	Features	Algorithm	Performance Metrics				
			Fitness	DR (%)	Accuracy (%)	FPR	C T(s)
1	41	CFA	74.455	71.087	73.267	17.685	0.28
		ACO-CFA	76.88	77.11	78.55	8.777	0.50
2	35	CFA	78.004	69.526	75.013	2.21488	0.25
		ACO-CFA	82.215	75.661	81.85	1.21	0.47
3	30	CFA	78.235	69.538	75.167	1.471	0.21
		ACO-CFA	84.212	77.66	83.77	1.01	0.43
4	25	CFA	83.623	78.212	81.714	3.752	0.19
		ACO-CFA	88.66	85.47	88.55	1.88	0.39
5	20	CFA	92.922	91.362	92.372	3.438	0.15
		ACO-CFA	95.314	95.44	96.11	1.64	0.36
6	15	CFA	93.070	91.500	92.500	3.372	0.13
		ACO-CFA	96.121	98.11	98.55	1.11	0.32
7	10	CFA	93.265	92.051	92.837	3.900	0.12
		ACO-CFA	97.211	96.12	96.32	1.81	0.29
8	5	CFA	95.524	91.000	91.986	3.917	0.10
		ACO-CFA	98.788	95.21	95.33	2.31	0.26

Pheromone value is updated locally for selected feature by using eq.2.

Until all ants are finished

5. CFA

Repeat

Initialize and keep the best solution and the average value of the best solution.

Use interaction operator between chromatophores and iridophores cells in case 1 and 2, to produce a new solution based on the reflection and the visibility of pattern.

Use iridophores cells operators in case 3 and 4 to calculate new solutions based on the reflected light coming from best solution and the visibility.

Use leucophores cells operator in case 5 to produce new solution by reflecting light from the area around the best solution and visibility of the pattern (local search).

Use leucophores cells operator in case 6 for random solution by reflecting incoming light (global search)

Until termination condition

The best subset is sent to ant.

6. Pheromone value is updated globally by using eq.3

7. l=l+1

8. The best feature subset is recorded.

6. Experimental Results

KDDCUP'99 dataset used for evaluation of ACO-CFA.The dataset contain 41 features. With 41 features the experiment is started to select the subset of relevant features. The experiment is done for various numbers of features. Fitness value is calculated using the formula

$$Fitness\ Function = \alpha * Detection\ Rate + \beta(1 - False\ Positive\ Rate)$$

$$Where\ \alpha and\ \beta\ are\ constants.$$

α represents the importance of Detection Rate

β represents the importance of False Positive Rate.

$$If\ Detection\ Rate\ is\ given\ more\ importance, \alpha = 0.7\ and\ \beta = 0.3$$

$$If\ Detection\ rate\ and\ False\ Positive\ Rate\ is\ considered\ as\ equally\ important, \propto = 0.5\ and\ \beta = 0.5$$

The metrics such as fitness, accuracy, detection rate, computation time and false positive rate is calculated. The experimental results are shown in table 1

The comparison of different metrics are shown in the graph

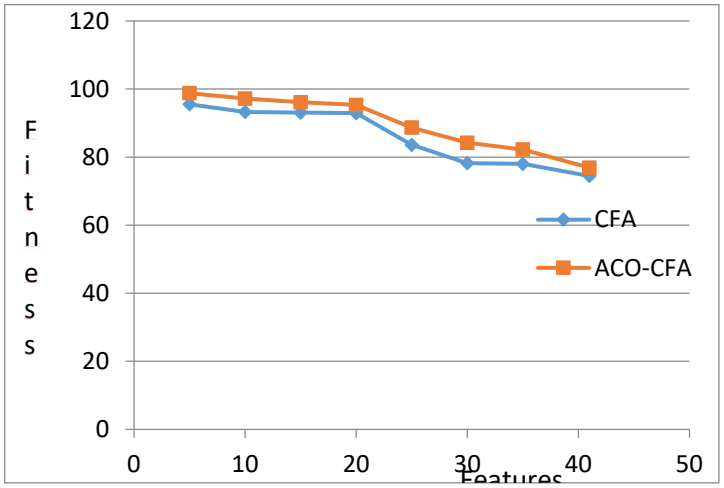


Figure 3 Comparison of fitness

The fitness of CFA and ACO-CFA is shown in figure3. Figure 4 shows the detection rate of CFA and ACO-CFA. The accuracy and false positive rate is shown in figure 5 and figure 6.

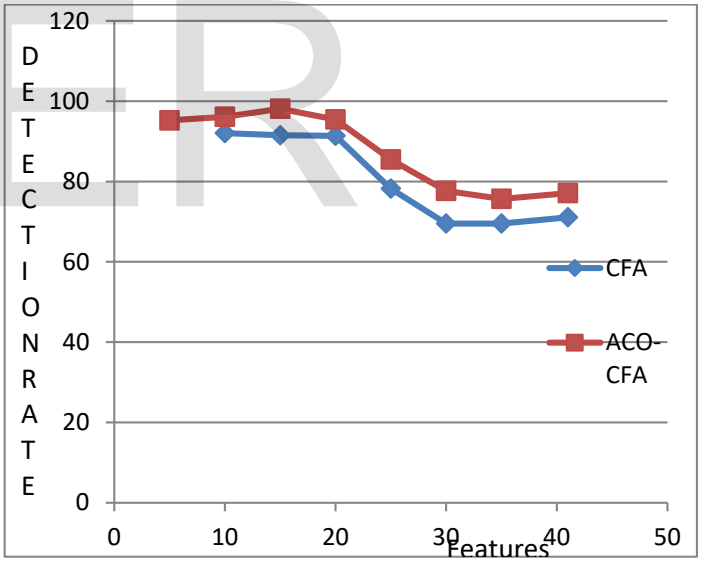


Figure 4 .Comparison of detection rate

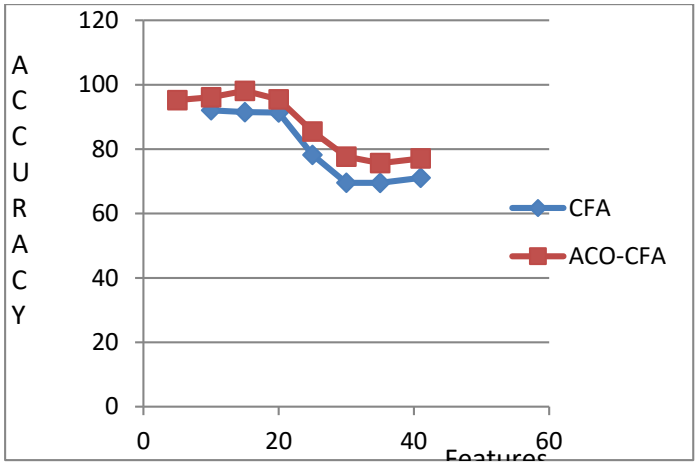


Figure 5 .Comparison of accuracy

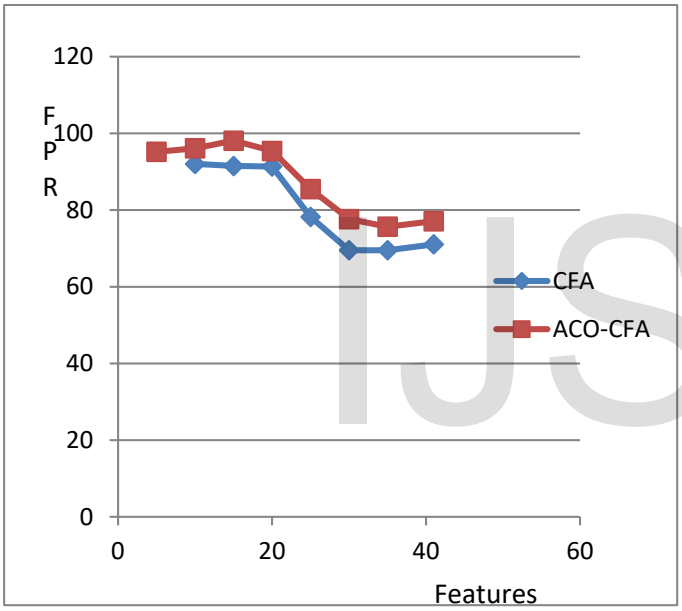


Figure 6.Comparison of false positive rate

6. Conclusion

ACO-CFA generates the initial solution and gives it to the CFA for feature selection. The overall performance has been improved when compared to CFA.ACO generates the optimal initial population. From the optimal initial solution, the optimal subset o features are selected for further classification. Decision tree classifier is used for evaluation. The detection rate is increased and the accuracy is also improved.ACO performs the global search in a better way.

Table 2 Comparison of different algorithms

S.No	Algorithms	Feature length	Detection rate (%)	Accuracy (%)	Computation Time(Sec)
1	Bat	-	92.94	-	0.43s
2	Bee	8	95.75	-	-
3	Bee-MC	41	89.11	95.6	-
4	PSO	12	-	94.49	-
5	Hybrid Kernel PCA	12	94.22	-	-
6	CFA	10	92.05	-	0.28
7	ACO-CFA	12	98.01	97.8	0.29

Since the initial solution is optimal, the final solution is also optimal.

Acknowledgments This work is supported by the University Grants Commission under Rajiv Gandhi National Fellowship for SC students.

Bibliography

[1] Rowland, Craig H, "Intrusion detection system ," *U.S. Patent No. 6,405,318*, 2002.

[2] Sun, Meng, and Tom Chen , "Network intrusion detection system ," *U.S. Patent Application No. 12/411,916*, 2010.

[3] De Boer, Pieter, and Martin Pels , "Host-based intrusion detection systems ," *Amsterdam University* , 2005.

[4] Kruegel, Christopher, and Thomas Toth, "Using decision trees to improve signature-based intrusion detection," *International Workshop on Recent Advances in Intrusion Detection*, 2003.

[5] Garcia-Teodoro, Pedro, et al., "Anomaly-based network intrusion detection: Techniques, systems and

challenges," *computers & security* , vol. 28, no. 1-2, pp. 18-28, 2009.

[6] Aslahi-Shahri et al. , "A hybrid method consisting of GA and SVM for intrusion detection system.," *Neural Computing and Applications*, pp. 1-8, 2015.

[7] Kolias, Constantinos, Georgios Kambourakis, and M. Maragoudakis, "Swarm intelligence in intrusion detection: A survey," *computers & security*, vol. 30, no. 8, pp. 625-642, 2011.

[8] Chung, Yuk Ying, and Noorhaniza Wahid, "A hybrid network intrusion detection system using simplified swarm optimization (SSO)," *Applied Soft Computing* , vol. 12, no. 9, pp. 3014-3022, 2012.

[9] Xiao, Lizhong, Zhiqing Shao, and Gang Liu, "K-means algorithm based on particle swarm optimization algorithm for anomaly intrusion detection," *Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on. Vol. 2. IEEE, 2006.*, vol. 2, pp. 5854-5858, 2006.

[10] Kuang, Fangjun, Weihong Xu, and Siyang Zhang, "A novel hybrid KPCA and SVM with GA model for intrusion detection," *Applied Soft Computing 18*, vol. 18, pp. 178-184, 2014.

[11] Ramos, Vitorino and Ajith Abraham, "Antids: Self orga nized ant-based c lustering model for intrusion det ection system," *Soft Computing as Transdisciplinary Science and Technology*, pp. 977-986, 2005.

[12] Kuang, Fangjun, et al. , "A novel SVM by combining kernel principal component analysis and improved chaotic particle swarm optimization for intrusion detection ," *Soft Computing* , vol. 19, no. 5, pp. 1187-1199, 2015.

[13] Hamid, Yasir, M. Sugumaran, and V. R. Balasaraswathi, "Ids using machine learning-current state of art and future directions ," *British Journal of Applied Science & Technology* , vol. 15, no. 3, 2016.

[14] Kazeem I. Rufai, Ravie Chandren Muniyandi, and Zulaiha A. Othman, "Improving Bee Algorithm Based Feature Selection in Intrusion Detection System Using Membrane Computing," *Journal of Networks*, vol. 9, no. 3, 2014.

[15] Mehmod, Tahir ,et al. , "Ant Colony Optimization and Feature Selection for Intrusion Detection," *Advances in Machine Learning and Signal Processing*, pp. 305-312, 2016.

[16] Mehdi Hosseinzadeh Aghdam, et al. , "Feature Selection for Intrusion Detection System using Ant Colony Optimisation," *International Journal of Network Security*, vol. 18, no. 3, pp. 420-432, 2016.

[17] Ghanem, Waheed Ali HM and Aman Jantan , "Novel Multi-Objective Artificial Bee Colony Optimization forWrapper Based Feature Selection in Intrusion Detection," *International Journal of Advance Soft Computing Applications*, vol. 8, no. 1, 2016.

[18] Chunying Cheng, Lanying Bao and Chunhua Bao, "Network Intrusion Detection with Bat Algorithm for Synchronization of Feature Selection and Support Vector Machines ," *International Symposium on Neural Networks.Springer International Publishing Switzerland*, pp. 401-408, 2016.

[19] Wang Xingzhu, "ACO and SVM Selection FeatureWeighting of Network ," *International Journal of Security and its Applications*, vol. 9, no. 4, pp. 129-270, 2015.

[20] Eesa, Adel Sabry, Zeynep Orman et al. , "A novel feature selection approach based on the cuttle algorithm for intrusion detection systems ," *Expert Systems with Applications* , vol. 42, no. 5, pp. 2670-2679, 2015.

[21] Dorigo, Marco and Christian Blum, "Ant colony optimization theory: A survey," *Theoretical computer science,2005,344 (2), 243-278.*, vol. 344, no. 2, pp. 243-278, 2005.