Support vector machine based Intrusion Detection System with Hybrid Swarm intelligence

Mrs. Charusheela P. Kamble[1], Prof. Mrs. D.P. Adhyapak[2]
[1, 2] Dept. of E&TC
[1, 2] PES's Modern College of Engineering, Pune, India

Abstract- *Intrusion Detection System (IDS) takes an important role in network security as it detects various types of attacks in the network. Swarm intelligence and machine learning techniques (SVM) were used to build different Intrusion Detection System. Support vector machine provides solutions for IDS. Support vector machines are supervised learning models with associated learning algorithm that analyze data used for classification. A hybrid algorithm is implemented to integrate Modified Artificial Bee Colony (MABC) with Enhanced Particle Swarm Optimization (EPSO) to predict the intrusion detection problem.*

The performance metrics like accuracy, sensitivity, false alarm rate, and training time are recorded for the intrusion detection dataset on applying the proposed MABC-EPSO classification algorithm.

Keywords- Intrusion Detection, SVM, PSO, ABC and NSL-KDD.

I. INTRODUCTION

Network security has become an indispensable factor of computer technology with the development of internet. The security of a computer system or network is compromised when an intrusion takes place. An intrusion can be defined as any set of actions that makes an attempt to compromise the integrity, confidentiality or availability of a resource. Intrusion prevention techniques such as firewalls, access control or encryption have failed to protect networks and systems from increasing attacks and malwares. As a result, Intrusion Detection System (IDS) have become an essential component of security infrastructure to detect these threats, identify and track the intruders. As IDS must have a high attack Detection Rate (DR), with a low False Alarm Rate (FAR) at the same time, construction of IDS is a challenging task.The main goal of IDSs is to identify and distinguish the normal and abnormal network connections in an accurate and quick manner which is considered as one of the main issues in intrusion detection system because of the large amount of attributes or features. To study about this aspect, data mining based network intrusion detection is widely used to identify how and where the intrusions occur.The main objective of work is (1) to select important features (2) to providea hybrid optimization algorithm based on Artificial Bee Colony (ABC) and Particle Swarm Optimization (PSO) algorithms for classifying intrusion detection dataset.ABC algorithm has powerful global search ability but poor local search ability, the PSO has powerful local search ability and poor global search ability.To provide a powerful global search capability and local search capability, hybridized MABC-EPSO is proposed which brings the two algorithms together so that the computation may benefit from both advantages. KDDCUP'99 intrusion detection dataset developed by MIT Lincoln Laboratory is used for experiments to find theaccuracy of the proposed hybrid approach.

II. LITERATURE REVIEW

Adriana-CristinaEnache1,VictorValeriu Patriciu2 [1] proposed an anomaly network intrusion detection approach using Information Gain for feature selection and Support Vector Machine optimized with Swarm Intelligence for classification. P. Amudha1,S. Karthik2, and S. Sivakumari1[2] proposed a hybrid algorithm which integrate Modified Artificial Bee Colony (MABC) with Enhanced Particle Swarm Optimization (EPSO) to solve the intrusion detection problem. Revathi and Malathi [3] provide a hybrid simplified swarm optimization to preprocess the data.

Satpute et al. [5] enhanced the process of intrusion detection system by combining PSO with machine learning techniques for the detection of anomaly in network intrusion detection system.In this paper, a hybrid algorithm based on ABC and PSO was proposed to classify the intrusion detection dataset.

The rest of this paper is organized as follows:Section III presents a background of the SVM algorithm. Section IV describes the PSO algorithm.Section V describes the ABC algorithm. Section VI describes the proposed model. Conclusion is presented in Section VII.

III. SUPPORT VECTOR MACHINE

Support Vector Machine is a trained classification algorithm oriented from statistical learning theory. SVM was developed by Vapnik. This method is utilized to classify classification problems having two distinct classes. SVM is used for many classification problems. In SVM, it's aimed to find the optimum hyper plane that separates two different classes. This optimum hyper plane divides the sample space so that distances of the different classes samples to the hyper plane are the most. The functions which are used to generate the hyper plane are called as kernel functions. These functions could be linear or non-linear. Linear, polynomial and radial basis function (RBF) kernels are most widely used kernel functions.

IV. PARTICLE SWARM

OPTIMIZATION (PSO)

PSO is a computation technique developed by Kennedy and Eberhart. The algorithm was inspired by bird flocking. This method optimizes a problem by trying to improve a candidate solution (the current location) with regard to a given measurement of quality (the fitness function) .It will search for the best parameters: C (cost) and σ based on the accuracy of the SVM algorithm.

PSO performs searches using a population (or swarm) of agents (called particles). Each particle i has a current position loci=(loci,1, loci, 2 ,….Loci,d)t and a current flying velocity Veli = (veli,1, veli,2 ,veli,d)t, where d is the problem dimension (d is two in our case). To discover the optimal solution, each particle moves in the direction of its previous best position (p_best) and its best global position (g_best) position.In PSO,a swarm consisting of N particles in a space of dimensional searches D. The i^thparticle is represented asX_i = (x_{i1}, x_{i2} , ...,x_{id}).The best location of any particle Pbest:Pi= (Pi1,pi2,……pid) and the velocity of the particle i is V_i=(V_{i1},V_{i2},V_{i3},…….V_{id})) The best overall particle in any swarm is indicated by Pg represents the particle's suitability. During each iteration, each particle updates the velocity according to the following equation

$$V^{t}_{id}=v^{t-1}id+C_1rand_1(P_{id}–x^{t-1}id)+C_2rand_2(P_{gd}-x^{t-1}id)$$

Where C1 and C2 indicate the acceleration coefficients d = 1,2, ... D and rand1 and rand2 random numbers are evenly distributed within (0, 1). Each particle moves to a new potential position as in the following equation,

$$x^{t}_{id=}x^{t-1}id+v^{t}\,id,\ \ d=1,2,……,D$$

V. ARTIFICIAL BEE COLONY

Artificial Bee Colony (ABC) is an algorithm based on the behavior of honey bee proposed by Karaboga and Basturk. The artificial bee colony consists of three groups: scout bees, observer bees, employed bees. In the context of optimization, the amount of food sources in the ABC algorithm represents the number of solutions in the population. The point of a good source of food indicates the position of a promising solution to the optimization problem. The four main steps of the procedure are as follows.

1. Initialization Phase: Scout bees randomly generate population size (SN) from food sources. The input vector containing variables is the food source which is the size of the search space of the objective function to be optimized, using the following equation initial food sources occur at random.

Xm=li +rand (0,1)*(ui-li)Where ui and li are the upper and lower limits of the objective function and solution space is a random number in the interval (0, 1).

2. Employed Bee Phase-e:The bees employed find a new source of food within the food source region; the employed bees reminiscent of a greater quantity of food source and share it with the bees spectators. The following equation determines the nearest food source and is calculated as, Vm i= xmi + Ømi (xmi- xki)

Where i is a randomized index parameter, xk is a randomly selected source of food, and Ømi is a random number within the range [-1, 1]. Fitness Food Sources, you need to find the optimal global solution is calculated,

Fiti =1/fi+1 ………fi >0

=1+|fi|………………fi<0 Where fi represents the objective value of the solution.

3. Onlooker Bee Phase:Observer Bees examines the effectiveness of food sources by observing the waggle dance in the dance region and then randomly selecting a rich source of food. The amount of a food source for its probability is evaluated using the equation,

Pi= fiti /∑ fiti

Where fiti indicates the suitability of the solution represented by the food source and the total number of food sources equal to the number of bees occupied.

4. Scout Phase: If the efficacy of a food source cannot be improved with the fixed number of trials, then scout bees remove solutions and try out new solutions at random.

- IDS model has proposed three different phases.
- **Data pre-processing:** In data pre-processing, data is prepared for classification and unused features and deleting duplicate instances.
- **Feature selection:** Two methods for selecting the feature are SFSM and RFSM methods.
- **Hybrid classification:** In hybrid classification, MABC-EPSO algorithm is utilized to improve the classification accuracy for the KDDCUP’99 dataset.

Data Pre-processing: The main objective of the pre-processing of data is to transform the raw data network into a suitable form for further analysis. If the dataset contains duplicate cases, classification algorithms consume more time and give inefficient results. This data set contains many redundant cases. The large amount of duplicate instances will make learning algorithms be partial towards the frequently occurring instances and will inhibit it from learning infrequent instances which are generally more unsafe to networks. Also, the existence of these duplicate instances will cause the evaluation results to be biased by the methods which have better detection rates on the frequently occurring instances. Eliminating duplicate instances helps in reducing false-positive rate for intrusion detection. Hence, duplicate instances are removed, so the classifiers will not be partial towards more frequently occurring instances. Moreover, irrelevant and redundant attributes of intrusion detection dataset may lead to complex intrusion detection model and reduce detection accuracy.Feature selection: As the data set is large, it is essential to eliminate insignificant features in order to distinguish normal traffic or intrusion in a timely manner. Feature subtypes are formed by the single feature method (SFSM), the random selection method (RFSM), and compare the two techniques. Feature selection methods reduce the characteristics of data sets that are intended to improve accuracy, reduce processing time, and improve efficiency to detect intrusions.

Hybrid Classification Approach: Artificial intelligence and machine learning techniques have been used to build different IDS, but showed the limits of achieving high detection accuracy and fast processing time. Computational intelligence techniques, known for their ability to adapt and to expose fault tolerance, high computational speed and resistance to noisy information, compensate for the limitations of these approaches. Hybrid algorithm combines the logic of both ABC and PSO.

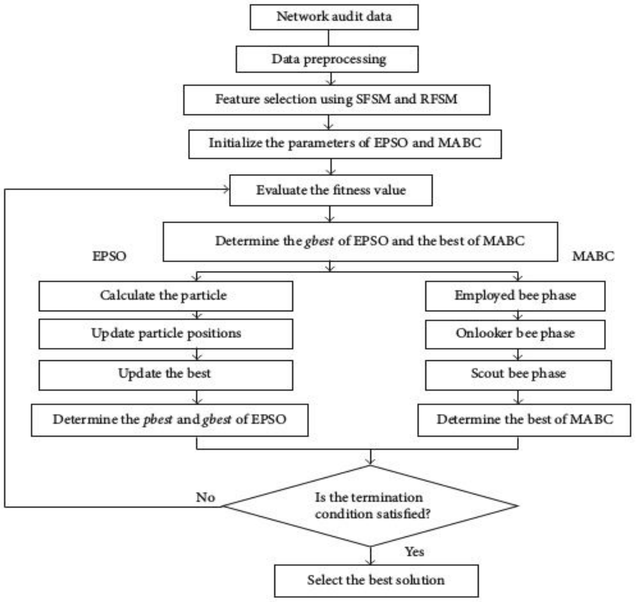
Network audit data
Data preprocessing
Feature selection using SFSM and RFSM
Initialize the parameters of EPSO and MABC
Evaluate the fitness value
Determine the *gbest* of EPSO and the best of MABC
EPSO
MABC
Calculate the particle
Update particle positions
Update the best
Determine the *pbest* and *gbest* of EPSO
Employed bee phase
Onlooker bee phase
Scout bee phase
Determine the best of MABC
No
Is the termination condition satisfied?
Yes
Select the best solution

Figure 1. Flowchart of the hybrid MABC-EPSO model

A new hybrid algorithm MACO-EPSO is proposed in project work. The MACO-EPSO has lowest training time amongst all classification algorithms.

Table 1. Accuracy (%) for classifiers

	NAÏVE BAYES	**SVM**	**MABC-EPSO**	**MACO-EPSO**
DOS	**82%**	**92%**	**99.837%**	**89.83%**
PROBE	**84%**	**95%**	**98.065%**	**88.06%**
U2R	**85%**	**96%**	**97.688%**	**87.68%**
R2L	**86%**	**96%**	**97.972%**	**87.97%**

Table 2. sensitivity (%) for classifiers

	NAÏVE BAYES	**SVM**	**MABC-EPSO**	**MACO-EPSO**
DOS	**83%**	**87%**	**99.858%**	**95.85%**
PROBE	**81%**	**80%**	**98.858%**	**95.85%**
U2R	**75%**	**79%**	**97.858%**	**95.85%**
R2L	**72%**	**81%**	**97.858%**	**95.85%**

Table 3. Training Time (in ms)

	Training Time (in ms)
SVM	**31ms**
MABC-EPSO	**15ms**
MACO-EPSO	**11ms**

VI. CONCLUSION

Intrusion Detection is a process of detection Intrusion in a computer system in order to increase the security. A hybrid algorithmic is proposed to integrate Modified Ant Colony (MACO) with Enhanced Particle Swarm Optimization (EPSO) to predict the intrusion detection problem.

The hybrid algorithm based on MABC and EPSO has highest detection rate and accuracy amongst the various algorithms proposed so far by various authors.The hybrid MABC-EPSO has accuracy up to 98.83% .A new hybrid approach MACO-EPSO is implemented.The hybrid MACO-EPSO has lowest training time amongst all classification algorithms.

REFERENCES

[1] Adriana-Cristina Enache1, VictorValeriu Patriciu2,"Intrusion Detection Based On Support Vector Machine Optimized with Swarm Intelligence" 9th IEEE International Symposium on Applied Computational Intelligence and Informatics • May 15-17, 2014 • Timisoara, Romania

[2] P. Amudha1, S. Karthik2, and S. Sivakumari1, A Hybrid Swarm intelligence Algorithm for Intrusion Detection Using Significant Features, The Scientific World Journal Volume 2015, Article ID 574589,15pages p://dx.doi.org/10.1155/2015/574589

[3] S. Revathi and A. Malathi, "Data preprocessing for intrusion detection system using swarm intelligence techniques," International Journal of Computer Applications, vol. 75,no. 6, pp. 22–27,2013

[4] A Detailed Analysis of the KDD CUP 99 Data ,Mahbod Tavallaee , Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbaniet , proceeding 2009 IEEE

[5] K. Satpute, S. Agrawal, J. Agrawal, and S. Sharma, "A survey on anomaly detection in network intrusion detection system using particle swarm optimization based machine learning techniques,"in Proceedings of the International Conference on Frontiers of Intelligent Computing:Theory and Applications (FICTA),vol. 199 of Advances in Intelligent Systems and Computing, pp.441–452, Springer, Berlin, Germany, 2013