Intrusion Detection using IP Binding in Real Network

Vishakha R. Deshmukh[1], Dr. Sheetal S. Dhande-Dandge[2]

[1]Student, Department of Computer Science & Engineering, SIPNA COET Amravati, Maharashtra, India
[2]Professor, Department of Computer Science & Engineering, SIPNA COET Amravati, Maharashtra, India

---***---

Abstract - *In the era of big data, with the increasing number of audit data features, human-centered smart intrusion detection system (IDS) performance is decreasing in training time and classification accuracy, and many SVM-based intrusion detection algorithms have been widely used to identify an intrusion quickly and accurately. So in the project to propose the IP binding technique in which the regular IP addresses in the network will be considered in which the regular authentication system will get workout with network evaluation for the determination of the intruder. The undefined IP determination technique is used to speed up the intruder detection system by implementing SVM verification with meta data verification. The SVM algorithm first optimizes the crossover probability and mutation probability of GA according to the population evolution algebra and fitness value then, it subsequently uses a feature selection method based on the genetic algorithm with an innovation in the fitness function that decreases the SVM error rate and increases the true positive rate with highly configured authentication technique. To perform evaluation mechanism in authentication of nodes and live performer evaluation and detection system implementation.*

Key Words**: Genetic Algorithm, Support Vector Machine, Intrusion Detection, IP Binding, Chromosome.**

1. INTRODUCTION

With the development and popularity of internet and network technologies the information security becoming more and more important. Compared with traditional network defence technology such as firewalls, human centered smart IDSs that can take initiative to intercept and warn of network intrusion has a great practical value. The question of how to improve effectiveness of smart network intrusion detection has become a focus of security [1]. Currently, there is much focus on the intrusion detection system (IDSs), which has close links for the safety of network service application [2]. The use of smart IDS is viewed as an effective solution for network security and protection against external threats. However existing IDS often has lower detection rate under new attacks and has high overhead when working with audit data and thus machine learning method have been widely applied in intrusion detection. To address this issue, several machine learning methods are have been extended. SVM, one of the machine learning technology, is a new algorithm based on statistical learning theory that has shows higher performance than the traditional learning method in solving the clarification problem of pattern recognition and speech recognition [3].

Compared with other classification algorithms, SVM can better solve the problems of small samples, non linearity and high dimensional. However, with the advent of era of big data, SVM encounters the problem of long training and testing times, high error rates and low true positive rates which limits the use of SVM in network intrusion detection. GA shows excellent global optimization ability via population search strategies and information exchange between individuals. Different from the traditional multi-point search algorithm. GA can easily avoid local optima.

In this review work GA and SVM are used to select optimal feature subset and optimize the SVM to improve the performance of the network intrusion detection system. GA and SVM algorithm enhance the effectiveness of the measures in detecting intrusion.

2. LITERATURE REVIEW

In the era of big data, intrusion detection becomes the most important topic in security infrastructure. To distinguish between attacks and normal network access, different machine learning methods are applied in IDS, including fuzzy logic[4], K nearest neighbour(KNN)[5], support vector machine(SVM), artificial neural network(ANN)[5], and artificial immune system(AIM) approaches[6]. Using IDS is an effective solution to provide security to the network against external threats. IDS is a control and protective measure that detects misused, abused, and unauthorized access to network resources.

SVM showed better performance than other traditional classification techniques. And SVM based IDS can improve IDS performance in term of detection rate and learning speed compare with traditional techniques. To address these problems, we use GA technology to supply fast and accurate optimization that can enable IDS to find optimal detection model based on SVM.

Genetic algorithm (GA) was proposed to improve the intrusion detection system (IDS) based on support vector machine(SVM)[10]. An intrusion detection method based on wavelet kernel least square was designed to improve the detection.

Capability of SVM in a complex non-linear system. However the training and testing time of the algorithm is relatively long. Genetic algorithm is dynamically adjust via a heuristic strategy, and the classification accuracy of the model is taken as a objective function to realize parameter optimization of the Gaussian kernel based SVM classification model.

Genetic operators are the key to optimization, and specifically, the crossover and mutation operators are used to maintain population diversity and avoid local optima. Currently, the crossover probability and mutation probability are constant during the period of population evolution and delay the convergence of the algorithm in the later evolution of the population, leading to the long training time of SVM. Therefore, the method proposed in the review paper changes the crossover probability and mutation probability of GA according to the evolutionary algebra and fitness value, which generates the population to speed up the search in the early evolution of the population and accelerates the convergence of the algorithm in the later evolution of the population.

Considering the abilities of GA and SVM algorithms used in the literature, in order to get better performance and improve the accuracy of the detection, this review work investigates the performance of GA and SVM classification method for feature selection.

3. DETAILS OF TOPIC

In the review work genetic algorithm (GA) and support vector machine (SVM) algorithms are used. GA is one of the most powerful tools to search in a large space with the potential to find the best solution in a search space. And SVM is a supervised machine learning algorithm which can be used for both classification and regression challenges. If GA is used to optimize the SVM based intrusion detection system, the training time is longer, and the error rate is higher when selecting the optimal feature subset. After selecting the optimal feature subset, the importance of the features is not sorted. For these reasons, the combination of GA and SVM is proposed.

3.1 Genetic Algorithm (GA)

The proposed GA based feature selection technique, in this section it describes proposed single objective optimization (SOO) based feature selection approach. It identifies the relevant sets of features for SVM based classifier.

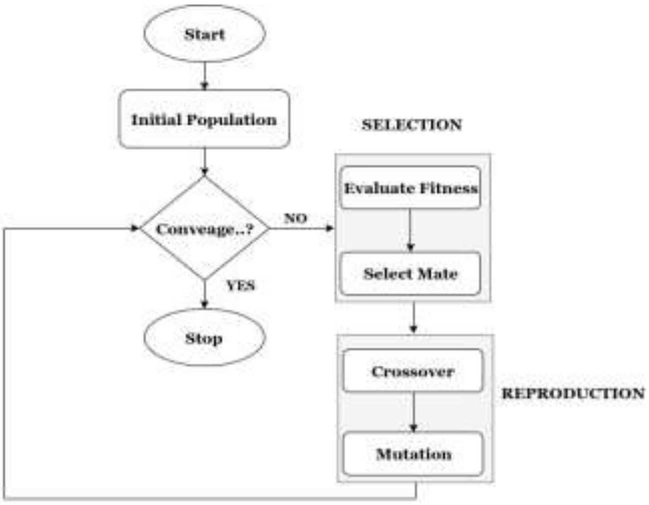

Fig -1: Working of Genetic Algorithm

3.1.1 Chromosome Representation and Population Initialization

If the total number of features is F, then the length of the chromosome is F.

Here, F = 12 (i.e., total 12 different features are available). The chromosome represents the use of 7 features, i.e., first, third, fourth, seventh, tenth, eleventh and twelfth for constructing the particular SVM based classifier. The entries of each chromosome are randomly initialized to either a 0 or 1. Here, if the ith position of a chromosome is 0 then it represents that ith feature does not participate in constructing the classifier. Else, if it is 1 then the ith feature participates in constructing the classifier. If the population size is P then all the P number of chromosomes of this population are initialized in the above way. The KDD CUP 99 data-set has 41 features in total. Thus the chromosome in our genetic algorithm is 41 bits long.

3.1.2 Fitness Computation

We use SVM for this purpose. For each subset considered, we take the 10 percent labeled KDD data set and used that as the test file. The remaining 90% data is used as the training set. The following steps are performed for each chromosome.

1. Suppose there are N number of features present in a particular chromosome (i.e., there are total N number of 1’s in that chromosome).

2. Construct the SVM based classifier with only these N features.

3. The SVM based classifier is tested on the test data.

4. The overall error rate of this SVM based classifier is calculated. The objective function corresponding to a particular chromosome is: f = error rate the objective is to minimize this objective function.

3.1.3 Genetic Operators

Roulette wheel selection is used to implement the proportional selection strategy. We use the normal single point crossover [9]. As an example, let the two chromosomes be: P1: 0 1 1 0 0 0 1 0 P2: 1 1 1 0 0 0 0 1 at first a crossover point has to be selected uniformly random between 1 to 8 (length of the chromosome) by generating some random number between 1 and 8. Let the crossover point, here, be 4. Then after crossover, the offspring's are: O1: 0 1 1 0 0 0 0 1 (taking the first 4 positions from P1 and rest from P2) O2: 1 1 1 0 0 0 1 0 (taking the first 4 positions from P2 and rest from P1) Crossover probability is kept equal to 0.9. Each chromosome undergoes mutation with a probability 0.2. Here, the value present at each position in a chromosome is flipped, i.e., if it initially contains 1 it will be replaced by 0; if it contains 0 it will be replaced by 1.

3.1.4 Termination Condition

In this approach, the processes of fitness computation, selection, crossover, and mutation are executed for a maximum number of generations. The best string seen up to the last generation provides the solution to the above classifier ensemble problem. Elitism is implemented at each generation by preserving the best string seen up to that generation in a location outside the population. Thus on termination, this location contains the best classifier ensemble.

3.2 Support Vector Machine (SVM)

“Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of feature you have) with the value of each feature being the value of a particular co-ordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well. The below fig. 2 shows the work of SVM.

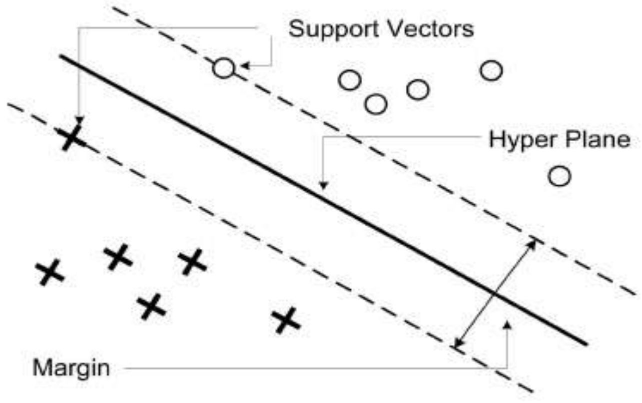

Fig 2- Working of Support Vector Machine

Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes i.e. Hyper-plane. Here, maximizing the distances between nearest data point (either class) and hyper-plane will help us to decide the right hyper-plane. This distance is called as Margin.

4. PROPOSED WORD AND OBJECTIVE

In this apropos mechanism will help to find the intruder detection by evaluation of IP binding technique in which the regular IP addresses in the network will be consider in which the regular authentication system will get workout with network evaluation for the determination of the intruder. The undefined IP determination technique is used to speed up the intruder detection system.

- To implement the SVM verification with meta data verification.
- To implement highly configure authentication technique.
- To improve the time barrier in SVM and GA based analytic.
- To perform evaluation mechanism in authentication of node.
- Live performer evaluation and detection system implementation.

5. SYSTEM ARCHITECTURE

This research is aimed at presenting an anomaly detection technique based on Genetic Algorithm (GA) and Support Vector Machine (SVM) function, which is to enhance the effectiveness of the measures in detecting intrusions. In the suggested solution, we use a hybrid model consisting of GA and SVM rather than primary algorithms (SVM). Using a hybrid model composed of GA and SVM is a very recent attempt, and today a large number of researchers are working on the combination of GA and SVM in order to improve the performance of classification for SVM.

As discussed by the researchers, hybrid learning methods would be an appropriate candidate for achieving the best possible accuracy and detection rate [14,15]. For more clarification, the combination of at least two learning approaches together is named hybrid learning. Combination of SVM and GA is known as a GA–SVM feature selection algorithm. Classification and regression problems would be solved by SVM. GA is a global optimal algorithm based on survival of the t test in Darwin’s theory of evolution. In order to select an optimal feature subset, the integration of two above mentioned methods can be useful

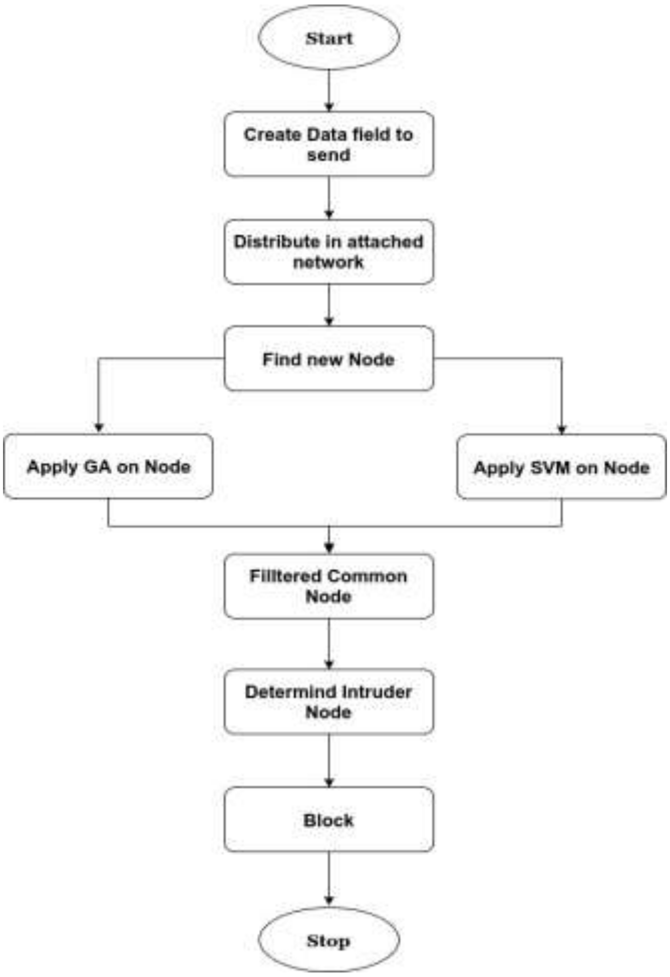

Fig -3: GA-SVM Flowchart

As shown in the Fig.4, are the steps that follows the to detect intruder in the network.

Step 1. It initialize when two nodes are communicating and are connected in the same network. Let us take Node A and Node B respectively.

Step 2. Node A Create a Data packet to share with Node B with a general Checksum that includes a hash value and some flags.

Step 3. Then Node A sends the data packet in the network so that it can be downloaded by the Partner Node (Node B in this case). If the Data packet gets modified or Flag Value is changed from 0 to 1 then we identify that there is an intruder (Node C) in our network.

Step 4. After detecting flag value change / Data packet modification our algorithm will looks for all the newly joined nodes in the Network.

Step 5. After getting the Initial Population, we will simultaneously run our hybrid algorithm (i.e. GA & SVM) on the initial population to get the optimized filtered common nodes as a combined output from our hybrid algorithm.

Step 6. The output of the GA-SVM algorithm will provide the filtered common nodes (i.e. Unauthorized Nodes.

Step 7. After getting the filtered Common node in the network algorithm will find for the intruder in the network.

Step 8. Once the intruder is detected it will be blocked.

6. CONCLUSION

Our project proposes an intrusion detection algorithm (SVM-GA) based on the genetic algorithm (GA) and uses support vector machine (SVM) algorithm. First, this project makes effective use of the GA population search strategy and the capability of information exchange between peer by optimizing the crossover probability and mutation probability of GA. A newly developed function is proposed that can minimize the SVM error rate and grow the true positive rate. Finally, the kernel parameter γ, the penalty parameter C and the feature weights are modified, and the working of SVM is improved. Our project results show that the improved intrusion detection technology based on the genetic algorithm (GA) and support vector machine (SVM) developed in this project.

ACKNOWLEDGEMENT

A moment of pause, to express deep gratitude to several individuals, without whom this project could not have been completed. We feel immense pleasure to express a deep sense of gratitude and indebtedness to our guide Dr. S. S. Dhande, for constant encouragement and noble guidance. We express our sincere thanks to Dr. V. K. Shandilya, Head of Department, Computer Science & Engineering, and the other staff members of the department for their kind co-operation.

We express our sincere thanks to Dr. S. M. Kherde Principal Sipna College of Engineering & Technology for his valuable guidance. We also express our sincere thanks to the library staff members of the college. Last but not least we are thankful to our friends and our parents whose best wishes are always with us.

REFERENCES

[1] Zhixin Sun, Peiying Tao, Zhe Sun, “An Improved Intrusion Detection Algorithm Based on GA and SVM.” DOI 10.1109/ACCESS.2018.2810198, IEEE Access.

[2] T. Yerong, S. Sai, X. Ke, and L. Zhe, “Intrusion detection based on support vector machine using heuristic genetic algorithm,” in Communication Systems and Network Technologies (CSNT), 2017 Fourth International Conference on. IEEE, 2014, pp. 681–684.

[3] Y. G. Liu, K. F. Chen, X. F. Liao, and W. Zhang, “A genetic clustering method for intrusion detection,” Pattern Recognition, Elsevier, pp. 927–942. 2004.

[4] W. M. Hu, W. Hu and S. Maybank,"AdaBoost-Based Algorithm for Network Intrusion Detection",IEEE IEEE Trans. Syst., Man, Cybern. B, Cybern, IEEE Computational Intelligence Society, pp.577-583. 2008

[5] I. Steinwart, D. Hush,C.Scovel, "A Classification Framework for Anomaly Detection", Journal of Machine Learning Research,pp.211-232 , 2005 .

[6] S. Ben-David, M. Lindenbaum; "Learning distributions by their density levels: a paradigm for learning without a teacher". Journal of Computer and System Sciences,Elsevier ,pp.171–182. 1997

[7] B. Scholkopf, R. Williamsonx ,"Support Vector Method for Novelty Detection", Journal of Machine Learning Research ,pp.582–588,2000

[8] Abraham A, Corchado E, Corchado JM (2009) Hybrid learning machines. Neurocomputing 72(13):2729–2730

[9] Alcala□R, Alcala□-Fdez J, Casillas J, Cordo□ O, Herrera F (2006) Hybrid learning models to get the interpretability–accuracy trade-off in fuzzy modeling. Soft Comput 10(9):717–734

[10] H. M. Lee, C. M. Chen, “A Self-Organizing HCMAC Neural-Network Classifier”, IEEE Trans. Neural Networks, IEEE Computational Intelligence Society, pp.15-27, 2003.

BIOGRAPHIES

Pursing M.E. Degree in Computer Science &Engineering, From Sipna College of Engineering and Technology, Amravati, Sant Gadge Baba Amravati University.

Dr. Sheetal S. Dhande-Dandge
Professor & Head, Deptt of Computer Science & Engineering, Sipna College of Engg and Technology, Amravati.Executive Member, CSI Amravati Chapter. MCSI, FIETE, MIE, MISTE.