ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 5, May 2014

Algorithm used in Intrusion Detection Systems: a Review

Harsimran kaur

Student, Department of computer science and engineering, S.G.G.S.W. University, Fatehgarh sahib, India.

ABSTRACT: -Communication through network should be safe and reliable with low maintenance cost. There is a process that used to detection activity of network that process is known as Network Intrusion Detection System (NIDS). IDS provide protection from attacker and detect activity. Modifications are applied on IDS time to time to make its performance better.There are someintrusion detection system approaches which we will discuss in this paper, like Bee Colony Optimization Algorithm (BCO), Ant Colony Algorithm (ACO) and Genetic Algorithm (GA). Main objective of this paper is to provide comparative study on these algorithms in data mining.

KEYWORDS: - Intrusion Detection System, Bee Colony Optimization, Ant Colony Optimization, Genetic Algorithm.

I. INRODUCTION

Now days, Network security infrastructure depend upon intrusion detection system. IDS that provide security from unknown intrusion attacks. There are many attacks are coming market day by day so organization should be ready to face that attacks and find out the proper way to handle that attacks. It is not possible to stop new attacks but we can handle that attacks. IDS are used as a defensive mechanism whose primary motive is to avoid the discontinuity in going work by considering all possible attacks on a system.

IDS are vast area for research in comparison to other existing areas. Main objective of this paper to study about existing algorithms and comparison of them on basis of advantages, disadvantages, and features etc. However, due to its mission critical nature, it has attracted significant attention towards itself. The main trend in this area is going about to find out better algorithm with the help of this we can get better results as compared to old technique used.

Intrusion detection is a process that can detect activity on both level host and network. There are to main ID techniques available that are anomaly detection andmisuse detection. Pattern for popular attacks created for matching with data and check for identity as intrusion or not [1]. Misuse detection model work like antivirus applications.IDS make so that it recognize normal activity and traffic recognize as attack. There are also many ways to accomplish this like we can use artificial intelligence techniques. We require data for testing and define patterns or rules for various processes. There is requirement of sensor for IDS. There sensor is system on which we can install and run IDS.

Traditional IDS model showing in Figure 1, here sensor machine is used to generate security events and to monitor and control that events there is management console. In the next sections we describe how algorithms (BCO, ACO) of IDS can be implemented.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 5, May 2014

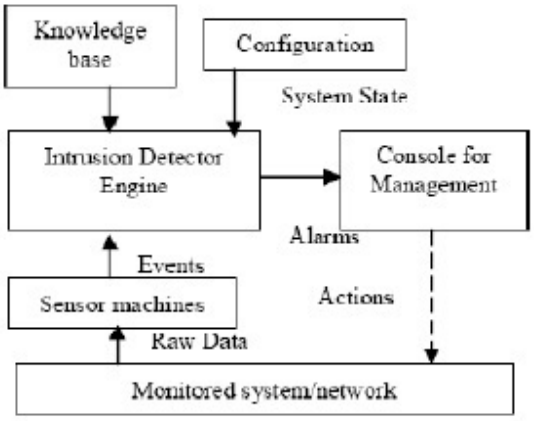

Fig. 1. Traditinal IDS Model[1]

II. BEE COLONY OPTIMIZATION ALGORITHM

a) Basic concept

Bees Algorithm is a searching algorithm on basis of population.Its concept is same as behavior of honey bee for gathering food.There are two optimization techniques combinational and continuous for which BCO is compatible.A colony of honey bee can extend up to a long distance of 15 km in search of food in any direction. A colony prospers by deploying its foragers to good fields.

In basic way, nectar or pollen of flowers exists from which food can access with less effort and in much amount. The area is very less visited where the flower of pollen or nectar is in small amount [2]. Scout bee start foraging process by patching searched flowers. These scout bees have random nature in moving from one patch to another. The bee colony is able to quickly switch the focus of the foraging effort on the most profitable flower patches [3].During the harvesting season, exploration of colony continues,more percentageof the population as scout bees. A patch that is found by scout bees are basis of certain quality threshold like sugar contents, when scout bees return to hive. After the threshold rating go to dance floor to perform a dance known as the waggle dance [4]. This dance makes colony communication and in that communication these bees collect three pieces of information about flower patch: a) in which direction those patch found b) its distance from hive c) its quality rating.

b) Algorithm:-

The Bees Algorithm is an optimization algorithm inspired by the natural foraging behavior of honey bees to find the optimal solution [5]. We have required to set number of parameters, like n- Number of scout, e- Number of best sites out of m selected sites, m- Number of sites selected out of n visited sites, (m-e)- Number of bees recruited for the other selected sites

Steps to follow for algorithm:

Step1. Initialize population with random solutions.
Step 2. Evaluate fitness of the population.
Step 3. While (stopping criterion not met) //Forming new population.
Step 4. Select sites for neighborhood search.
Step 5. Recruit bees for selected sites (more bees for best e sites) and evaluate fit nesses.
Step 6. Select the fittest bee from each patch.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 5, May 2014

Step 7. Assign remaining bees to search randomly and evaluate their fit nesses.
Step 8. End While.

III. ANT COLONY OPTIMIZATION ALGORITHM

A. Basic concepts

In real world, ants wander randomly, and after getting food return to their colony while laying down trails. After that if other ants find such a path, then this path like to follow that trail, not randomly. If those ants find food then the trail starts evaporate and this will effect on the strength of that trail.Pheromone evaporation also has the advantage of avoiding the convergence to a locally optimal solution. If there were no evaporation at all, the paths chosen by the first ants would tend to be excessively attractive to the following ones. In that case, the exploration of the solution space would be constrained. Thus, when one ant finds a good (i.e., short) path from the colony to a food source, other ants are more likely to follow that path [2].

B. Algorithm:-

We have to follow a procedure as following
procedureAnt colony optimization
Set Initialize parameters, pheromone trails
while(termination condition not met)
do
Construct Ant Solution
Update Pheromone Trails
Daemon Actions
end
end

IV. RELATED WORK

[21] A Borji has presented a unique technique of four classifiers presented namely ANN , SVM , Knn and Decision tree for the purpose of the intrusion detection . According to him , the intrusion is the additional data in the predefined data set and some classifier is required to analyse the data set if there is additional data . His work would help the future research workers in understanding the exact concept of the intrusion detection system .
[22] **Y Liao** has extended the work done by [1] and has used the K nearest algorithm as a classifier . He has also tried to change the basic algorithm structure presented in [1] and his results are efficient . The work done in [1] can be modified further by adding some some more efficient classifier or optimization algorithm or technique .
[23] **S. Jha** has kept HMM as their basic training method in terms of detection of the intrusion in the network . Furthermore they have also used classifier like SVM to detect the intrusion system . The problem of the SVM is the way it takes the data as an input . There are several other better classifiers which are available in this region and they can be used .
[24] **T Lappas** have explained all the ways of detecting an intrusion in a network . The main aim of his work is to highlight all the methods present for an intrusion system. He has explained SVM , nearest neighbor and decision tree in detail . His provided information will help out students in the further development .

[25] **H. A .Nguyen** have used the classifier model as a selection algorithm of the data. Although his work is quite efficient and can be referred further also but using the classifier algorithm as an optimization algorithm will not be suitable .

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 5, May 2014

V. RESULT COMPARISON OF BCO AND ACO

Figure 2. Shows if the first process of bee colony and second process of ant colony is considered then the processing time (Milliseconds) of ant colony is less than the bee colony optimization. The second comparison graph is shown in figure. 3.

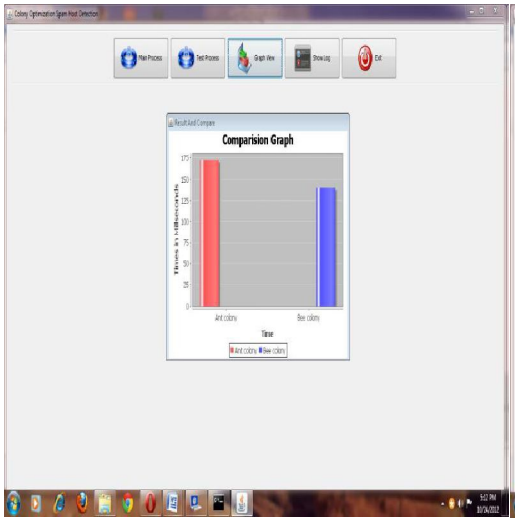

Fig.2 The first comparison graph

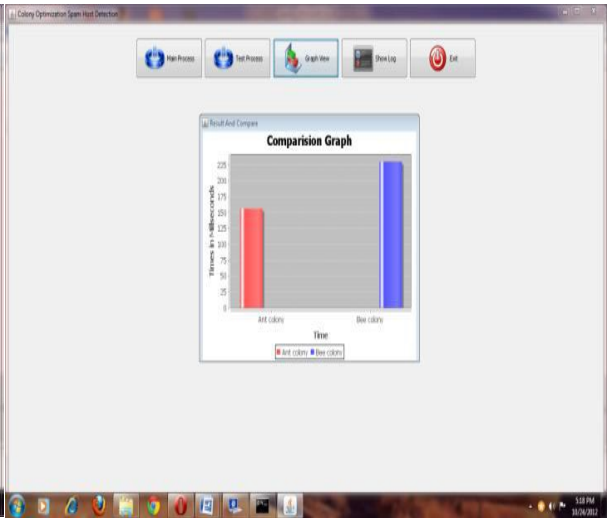

Fig.3. The second comparison graph

VI. GENETIC ALGORITHM IN IDS

a) Basic Concept

A Genetic Algorithm (GA) is a programming technique that mimics biological evolution as a problem-solving strategy [7]. It is based on Darwinian's principle of evolution and survival of fittest to optimize a population of candidate solutions towards a predefined fitness. GA uses an evolution and natural selection that uses a chromosome-like data structure and evolve the chromosomes using selection, recombination and mutation operators. The process usually begins with randomly generated population of chromosomes, which represent all possible solution of a problem that are considered candidate solutions.

b) Algorithm :

Initialize chromosomes for comparison
Input : Network audit data (for training)
Output : A set of chromosomes
1. Range = 0.125
2. For each training data
3. If it has neighboring chromosome within Range
4. Merge it with the nearest chromosome
5. Else
6. Create new chromosome with it
7. End if

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 5, May 2014

8. End for

VII. BAYESIAN NETWORKS

Uncertain information is handling by a graphical model that is known as Bayesian network [8, 9]. Two components of Bayesian network are as following:

- A graphical component of a directed acyclic graph (DAG) where events are represented by vertices and relationship between these events by edges.
- Numerical components consisting in quantification of different links in DAG by conditional probabilistic distribution of each node in contexts of its parents.

Simple network of Bayesian [10] is created with a parent node and other are children nodes with compose of DAG, parent node is unobserved node and children nodes are observed.If discuss about classification of data then Bayesian is very suitable algorithm for that purpose. It properly deals with problem of classification [11]. Relationships between some variables are encoded by a Bayesian network. Statistical method is combined with this network to detect intrusion with many advantages [12]. This network has capability to encoding between interdependencies between variables. Main disadvantage of this network is that its results are same as threshold based system but high level effort required in Bayesian network for computation as compared to threshold based system [13].

VIII. MARKOV MODELS

In this section, we discuss two main approaches like Markov chains and Markov models. When states are interconnected through some transition probabilities make a set that is known as Markov chain, due to which capability of model can be determined. Firstly probability is estimated on basis of normal behavior of target system during first phase. The detectionof anomalies is then carried out by comparing the anomalyscore (associated probability) obtained for the observedsequences with a fixed threshold.In the case of a hidden Markov model, the system ofinterest is assumed to be a Markov process in which states andtransitions are hidden. Only the so-called productions areobservable.Markov-based techniques have been extensively used inthe context of host IDS, normally applied to system calls [14]. In network IDS, the inspection ofpackets has led to the use of Markov models in someapproaches [15, 16]. In all cases, the model derived for the target system hasprovided a good approach for the claimed profile, while, as inBayesian networks, the results are highly dependent on theassumptions about the behavior accepted for the system.

IX. NEURAL NETWORKS

With the aim of simulating the operation of the human brain(featuring the existence of neurons and of synapses amongthem), neural networks have been adopted in the field ofAnomaly intrusion detection, mainly because of their flexibilityAnd adaptability to environmental changes. This detectionApproach has been employed to create user profiles [17], to predict the next command from a sequence ofPrevious ones [18], to identify the intrusiveBehavior of traffic patterns [19], etc.However, a common characteristic in the proposed variants,from recurrent neural networks to self-organizing maps [20], is that they do not provide a descriptiveModel that explains why a particular detection decision hasbeen taken.

X. TABLE FOR DISCUSSED ALGORITHMS

Algorithm Name	Advantage	Disadvantage
BEE COLONY OPTIMIZATION	• The algorithm has local search and global search ability,	• Random initialization, • The algorithm has several

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 5, May 2014

ALGORITHM	• Implemented with several optimization problems, • Easy to use, • Available for hybridization combination with other algorithms.	parameters, • Parameters need to be tuned.
ANT COLONY OPTIMIZATION ALGORITHM	• The algorithm has strength in both local and global searches. • Implemented with several optimization problems.	• Random initialization, • The algorithm has several parameters, • Parameters need to be tuned, • Probabilistic approach in the local search.
GENETIC ALGORITHM IN IDS	• Chance of being selected of a solution as a parent depends on the fitness value of that solution • It can solve every optimization problem which can be described with the chromosome encoding. • It solves problems with multiple solutions. • Since the genetic algorithm execution technique is not dependent on the error surface, we can solve multi-dimensional, non-differential, non-continuous, and even non-parametrical problems. • Structural genetic algorithm gives us the possibility to solve the solution structure and solution parameter problems at the same time by means of genetic algorithm. • Genetic algorithm is a method which is very easy to understand and it practically does not demand the knowledge of mathematics. • Genetic algorithms are easily transferred to existing simulations and models.	• Certain optimization problems (they are called variant problems) cannot be solved by means of genetic algorithms. This occurs due to poorly known fitness functions which generate bad chromosome blocks in spite of the fact that only good chromosome blocks cross-over. • There is no absolute assurance that a genetic algorithm will find a global optimum. It happens very often when the populations have a lot of subjects. • Like other artificial intelligence techniques, the genetic algorithm cannot assure constant optimization response times. Even more, the difference between the shortest and the longest optimization response time is much larger than with conventional gradient methods. This unfortunate genetic algorithm property limits the genetic algorithms' use in real time applications. • Genetic algorithm applications in controls which are performed in real time are limited because of random solutions and convergence, in other words this means that the entire population is improving, but this could not be said for an individual within this population. Therefore, it is unreasonable to use genetic

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 5, May 2014

		algorithms for on-line controls in real systems without testing them first on a simulation model.
BAYESIAN NETWORKS	• **Interpolation** Bayesian learning methods interpolate all the way to pure engineering. • An associated language for specifying priors and posteriors • Bayesian learning involves specifying a prior and integration	• Information theoretically infeasible • Computationally infeasible • Un-automatic
MARKOV MODELS	• Markov models are relatively easy to derive (or infer) from successional data. • The Markov model does not require deep insight into the mechanisms of dynamic change, but it can help to indicate areas where such insight would be valuable and hence act as both a guide and stimulant to further research. • The basic transition matrix summarizes the essential parameters of dynamic change in a way that is achieved by very few other types of model. • The results of the analysis of Markov models are readily adaptable to graphical presentation, and, in this form, are frequently more readily presented to, and understood by, resource managers and decision-makers. • The computational requirements of Markov models are modest, and can easily be met by small computers, or, for small numbers of states, by simple calculators.	• The lack of dependence on functional mechanisms reduces their appeal to the functionally orientated ecologist. • Departure from the simple assumptions of stationary, first-order Markov chains while, conceptually possible, makes for disproportionate degrees of difficulty in analysis and computation. • In some areas, the data available will be insufficient to estimate reliable probability or transfer rates, especially for rare transitions. For example, it may not be possible to observe sufficient transitions from a given transient set of states to a closed state where this transition is dependent on a rare climatic event, even though the value of this parameter is of vital importance in the dynamics of the community. • Like all other successional models, the validation of Markov models depends on predictions of system behavior over time, and is therefore frequently difficult, and may even be impossible, for really long periods of time.
NEURAL NETWORKS	• A neural network can perform tasks that a linear program cannot. • When an element of the neural network fails, it can continue without any	• The neural network needs training to operate. • The architecture of a neural network is different from the

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 5, May 2014

	problem by their parallel nature. • A neural network learns and does not need to be reprogrammed. • It can be implemented in any application and without any problem.	architecture of microprocessors therefore needs to be emulated. • Requires high processing time for large neural networks.

XI. CONCLUSION

With the comparison of both algorithm BCO and ACO we survey that by using GA we can also improve results by following the procedure as discuss in this paper. The parameter discuss in above algorithms are used to modified in GA. Each algorithm try to do best in a particular way but there are always some limitations that provide option for researcher to deign better algorithm than existing. Keep in mind the problems of existing algorithms we have to decide to follow the genetic algorithm due to which our processing time (milliseconds) will improve as compare to existing algorithms.

REFRENCES

1. Vivek K. Kshirsagar, Sonali M. Tidke& Swati Vishnu**," Intrusion Detection System using Genetic Algorithm and Data Mining: An Overview"** International Journal of Computer Science and Informatics ISSN (PRINT): 2231 –5292, Vol-1, Iss-4, 2012.
2. R. Sagayam, Mrs. K. Akilandeswari," **Comparison of Ant Colony and Bee Colony Optimization for Spam Host Detection**", International Journal of Engineering Research and Development eISSN : 2278-067X, pISSN : 2278-800X, www.ijerd.com Volume 4, Issue 8 (November 2012), PP. 26-32.
3. Tereshko V., Loengarov A. **,"Collective Decision-Making in Honey Bee Foraging Dynamics"**. Journal of Computing and Information Systems, 9(3), 1-7, 2005.
4. Von Frisch K. Bees: Their Vision, **"Chemical Senses and Language".** (Revised edn) Cornell University Press, N.Y., Ithaca, 1996.
5. Pham D.T., Ghanbarzadeh A., Koç E., Otri S., Rahim S., and M.Zaidi**"The Bees Algorithm – A Novel Tool for Complex Optimization Problems"**, Proceedings of IPROMS 2006 Conference, pp.454–461.
6. Mohammad SazzadulHoque, Md. Abdul Mukit and Md. Abu NaserBikas**," An Implementation Of Intrusion Detection System Using Genetic Algorithm"**, International Journal of Network Security & Its Applications (IJNSA), Vol.4, No.2, March 2012.
7. V. Bobor, **"Efficient Intrusion Detection System Architecture Based on Neural Networks andGenetic Algorithms",** Department of Computer and Systems Sciences, Stockholm University /Royal Institute of Technology, KTH/DSV, 2006.
8. F. V. Jensen, **"Introduction to Bayesian networks",** UCL press, university college, London, 1996.
9. J. Pearl, **"Probabilistic reasoning in intelligent system: networks of plausible interference",** Morgan Kaufmman, San Francisco (California).
10. P. Langley, W. Iba. , and K. Thompson, **"Decision making using probabilistic inference methods",** in proceeding of eight conference on uncertainty in artificial intelligence (UAI'92), pages 399-406, San Mateo, CA, 1992.
11. N. Friedman and M. Goldszmidt, **"Building classifiers using Bayesian networks"**, in proceeding of American association for artificial intelligence conference (AAAP'96), Portland, Oregon, 1996.
12. Heckerman D. **"A tutorial on learning with Bayesian networks".**Microsoft Research; 1995. Technical Report MSRTR-95-06.
13. Kruegel C., Mutz D., Robertson W., Valeur F. **"Bayesian eventclassification for intrusion detection"**. In: Proceedings of the19th Annual Computer Security Applications Conference; 2003.
14. Yeung DY, Ding Y. "**Host-based intrusion detection using dynamicand static behavioral models"**. Pattern Recognition 2003; 36(1): 229–43.
15. Mahoney M.V., Chan P.K. **"Learning nonstationary modelsof normal network traffic for detecting novel attacks"**.In: Proceedings of the Eighth ACM SIGKDD; 2002. p. 376–85.
16. Este´vez-Tapiador J.M., Garcı´a-Teodoro P., Dı´az-Verdejo J.E.**"Detection of web-based attacks through Markovian protocol parsing"**. In: Proc. ISCC05; 2005 p. 457–62.
17. Fox K., Henning R., Reed J., Simonian, R**. "A neural networkapproach towards intrusion detection"**. In: 13th National Computer Security Conference; 1990. p. 125–34.
18. Debar H., Becker M., Siboni, D. **"A neural network componentfor an intrusion detection system".** Symposiumon Research in Computer Security and Privacy; 1992. p. 240–50.
19. Cansian A.M., Moreira E., Carvalho A., Bonifacio J.M. **"Networkintrusion detection using neural networks".** In: InternationalConference on Computational Intelligence and MultimediaApplications (ICCMA'97); 1997. p. 276–80.
20. Ramadas M, Ostermann S, Tjaden B. **"Detecting anomalousnetwork traffic with self-organizing maps. In: Recent advances in intrusion detection, RAID"**. Lecture notes in computer science (LNCS), vol. 2820; 2003. p. 36–54.
21. Ali Borji ," **Combining Heterogeneous Classifiers for Network Intrusion Detection "**© Springer-Verlag Berlin Heidelberg 2007

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 5, May 2014

22. Y Liao , " Using K nearest Classifier for Intrusion Detection " ,IEEE 2010
23 .S. Jha , **Markov Chains, Classifiers, and Intrusion Detection , IEEE 2010 pp -257-311**
24. T Lappas , "Data Mining Techniques for (Network) Intrusion Detection Systems " IEEE 2011 pp-521-626 VOL 1
25 H. A .Nguyen and D Choi **Application of Data Mining to Network Intrusion Detection: Classifier Selection Model** Y. Ma, D. Choi, and S. Ata (Eds.): APNOMS 2008, LNCS 5297, pp. 399–408, 2008. © Springer-Verlag Berlin Heidelberg 2008