A Co-operative Game Theoretic Approach to Improve the Intrusion Detection System in a Network using Ant Colony Clustering

D.P.Jeyepalan
Research Scholar,
School of Computer Science, Engineering and Applications,
Bharathidasan University, Tiruchirappalli,
Tamilnadu, India

E.Kirubakaran
SSTP (Systems),
Bharat Heavy Electricals Ltd,,
Tiruchirappalli, India

ABSTRACT
Making a network foolproof is a very important task that every Intrusion Detection System should provide to the network. Areas of deployment of an IDS is also an important task that helps in efficient functioning of the system. Deploying the IDS in all the systems is a very inefficient strategy that reduces the performance of the entire network, while deploying the IDS in inappropriate or inefficient nodes leads to the system becoming vulnerable to attacks. The current proposal deals with improving the effectiveness of an Intrusion Detection System by selecting the appropriate candidates in the network. The candidacy selection is performed by initially clustering the nodes using Ant Colony Clustering algorithm and using Game Theoretical approaches for selection of heads and monitoring the environment for changes.

Keywords
Intrusion Detection; Ant Colony Clustering; Game Theory; Clustering

1. INTRODUCTION
The initial mode that explains the behaviour of ants which is commonly known as BM (Basic Model), was proposed by Deneubourg et al [1] . Lumer and Faieta[2] improved this model and proposed a new model called LF (Lumer and Faieta's Model). Both these models separate ants from the being clustered data objects, which increases the amount of data arrays to be processed. Both the algorithms use large amount of information, hence require large amount of memory space. Further, data analysis in both models is done by the ants and not automatically. Hence, it results in an increase in the computation burden. Further, the issue of outliers in the dataset is not dealt with, hence will lead to infinite movement. Based on the phenomenon that ants tend to group with fellows with similar features, the ants sleeping model (ASM) has been proposed. In this model, the ants directly represent the clustered data objects. The ants move according to the fitness of the surrounding environment and form into groups eventually and hence the corresponding data items are clustered.

Fuzzy ants[5] and clustering projected a method Fuzzy c-means and hard c-means for reformulated the fuzzy partition validity metric and to clustering the data. Clustering [15,16] web search results using fuzzy ants anticipated a method ACO, which was simulated using fuzzy IF-THEN rules or fuzzy logic. Clustering web search results have to be efficient and robust and the basic problem to be dealt with this is the absence of a priori information. Fuzzy controller design by clustering-aided ant colony optimization proposed a method Ant colony optimization (ACO) algorithm[7,9] (CACO) for improving both the design efficiency of a fuzzy controllers and its performance. Fuzzy ant based clustering proposed an optimization technique along with the usage of fuzzy if–then rules for clustering the data with no initial partitioning and known number number of clusters. An effective clustering algorithm with ant colony a method namely SACC algorithm and Jaccard index for solving unsupervised clustering problem and to identify the optimal cluster number. DWC [13,14] shows global connectivity and local compactness of the data. It first, calculates the set $\emptyset_i = \{\emptyset_{ij} \in X, j = 1, \dots L\}$ of the data object $x_i (i = 1, \dots N)_{L\,(L>0)}$ nearest neighbor [3], using the Euclidean distance formula. Then, link xi and (i= 1,…,N) L (L> 0) , and the link is undirected. Using this, a undirected graph G = (X,V) is created, and the adjacency $matrix^{R=[R_y]_{N.N}}$ of the data set, where:

$$R_{ij} = \{^{1, x_j \in \emptyset_i or x_i \in \emptyset_i}_{0, otherwise} \quad \text{.....(1)}$$

V is the set of all links between data. Rs = [Rij ^{s}]N × N , Rij^{s} is the number of steps, reachability paths between xi and xj .

The connectivity between data object xi and xj is defined as:

$$Conn(x_i, x_j) = \sum_{s-1}^{step} Conn^s(x_i, x_j) \quad \text{.....(2)}$$

$$Conn(x_i, x_j) = \begin{cases} \frac{log_L R_{ij^s}}{s-1}, if\ s > 1 \text{^} R_{ij^s > 1 \text{^} i \neq j} \\ 1\ if\ s = 1 \text{^} R_{ij^s > 0 \text{^} i \neq j} \\ 0, otherwise \end{cases} \quad \text{.....(3)}$$

The number of reachability paths between two points x_i and x_j determines the degree of connectivity between data point xi and xj, which automatically reflects the similarity between xi and xj. Hence, DWC (xi, xj)_ 1/Conn (xi, xj). We define the DWC distance of data objects as follows.

The DWC distance between data object xi and xj is defined as:

$$\text{Conn}\,(x_i, x_j) = \begin{cases} \frac{log_L R_{ij^s}}{s-1}, if\ s > 1 \text{^} R_{ij^s > 1 \text{^} i \neq j} \\ 1\ if\ s = 1 \text{^} R_{ij^s > 0 \text{^} i \neq j} \\ 0, otherwise \end{cases} \quad \text{…(4)}$$

$$\text{Dis}(x_i, x_j) = \sqrt{\sum_{v=1}^{m} \left| x_{iv} - x_{jv} \right|^2} \text{.....(5)}$$

It is the Euclidean distance between xi and xj, m denotes the number of the data object attributes,. If Dis(x_i, x_j) is short and Conn(x_i, x_j) is high. Then,the data object x_i,x_j has the highest probability of being in the same cluster. If Dis (x_i, x_j) is short but Conn (x_i, x_j) is low, Dis(x_i, x_j) will still be large, then, object xi and xj will not be grouped in the same cluster.

2. OUR APPROACH

The Game based approach for IDS enhancement is performed in three phases. The initial phase performs the process of clustering using the meta heuristic Ant Colony Clustering algorithm. The second phase helps in deciding the cluster heads that serves as the stations for running the IDS[10,11]. A cluster head is selected for every cluster, which runs the IDS. The final phase is the monitoring phase, that monitors the activity of the clusters and takes appropriate actions in case of any problems. If the cluster head in any cluster does not perform the monitoring to the required extent, then they are replaced with other efficient nodes.

2.1. Ant Colony Clustering

Ant Colony Clustering[4] is a nature inspired heuristic based algorithm that works on the basis of the behavior of ants in a colony while grouping food particles or their larvae. In the current paper, node distance is considered as the base property for the clustering process. The network is considered as a 2 dimensional grid[6] on which the nodes are placed. Every node is considered as an object and the similarity mapping is performed using the distance measure. A threshold[12] is set by the user, which determines if the current node is to be added to the cluster or not. Threshold is calculated using the following equation:

$$threshold = +(d_{max} - d_{min}) \times ClusterPrecision \quad \ldots(6)$$

Where d_{min} and d_{max} are the minimum and the maximum distances in the network, and the *ClusterPrecision* is determined by the administrator.

A random node is selected from the grid. The distance measure of the node is compared with the threshold to determine if the particular node belongs to the current cluster or is to be taken out. This process is repeated until all the nodes in the grid occupy their places in their corresponding clusters.

The probability of adding and dropping an item is defined by,

$$P_p(O_i) = \left[\frac{k_1}{k_1 + f(o_i)}\right]^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(7)$$

$$P_d(O_i) = \begin{cases} 2f(o_i) where\ f(o_i) < k_2 \\ 1, when\ f(o_i) \geq k_2 \end{cases} \ldots\ldots\ldots\ldots\ldots(8)$$

where :

- F(O_i) is a measure of the average similarity between two objects (O_i and O_j).
- d(O_i, O_j) is the dissimilarity between (O_i, O_j).
- $\propto$ determines the scale for dissimilarity.
- k_1 and k_2 are constants

2.2. Cluster Head Selection using Game Theory

Selection of a Cluster head has the biggest impact on an IDS. Since the IDS is generally deployed on the cluster heads, the cluster heads are expected to exhibit higher performance and high coverage area. Further, some nodes in a network are usually working for their own jobs other than the jobs for other nodes. Such nodes are to be identified and should be detained from being the cluster heads. The utilization of Game Theory[8] for this process is found to be very effective. Every node is aware of its neighbor nodes. The following describes the cluster head selection strategy.

In the first step, node $n_i \in [1,n]$ sends a Begin-Election message to all nodes $cluster^A$.The node identity, hash value, cost and time stamp are added to the message. This time stamp helps to avoid replay attacks. The hash function is used to avoid nodes from cheating and delivering a fake C_i as shown after. Then the election message is sent using the same parameters. Nodes that did not contribute in sending the Begin-Election message will be excluded from cluster's services. On receiving the Election messages, each node n_i verifies each received message with its corresponding hash value that has been sent in Begin-Election message. After the verification is accomplished, each node computes the SCF, which is the minimum valuation of cost of analysis. In the next step, if the leader is different from ni then it sends a Done-Election message to inform the leader that he has been elected. The elected leader forward a Confirm Leadership message indicating its acceptance. In the final step, the leader sets a timer T_2 then starts to verify the origin of all the Done-Election messages. If the timer expires without receiving all the Done-Election messages, then nodes who did not participate are excluded from the cluster's services. After the leader has been selected by all the nodes, all contributing nodes will be added to the protected list. To ensure fairness, a re-election timer is set, after whose expiration, re-election is performed.

2.3. Monitoring

Finally, selfish nodes might misbehave after election, which can be recognized by setting random checkers to ensure a catch-and-punish scheme in order to motivate an elected node to be faithful during the detection process. The selected checker is assumed to be cooperative since the benefit of the intrusion detection service dominates resources consumption. This is because we assume that the elected checkers mirror a portion of the computation done at the elected node which have a marginal effect on resource consumption.

3. RESULTS AND DISCUSSION

3.1. Simulation setup

The process was carried out with the Stanford dataset p2p-Gnutella04. This dataset was taken from Directed Gnutella P2P network from August 4 2002.

Table 1: Dataset Statistics

Dataset statistics	
Nodes	10876
Edges	39994
Average clustering coefficient	0.0062
Number of triangles	934
Fraction of closed triangles	0.001807
Diameter (longest shortest path)	9
90-percentile effective diameter	5.4

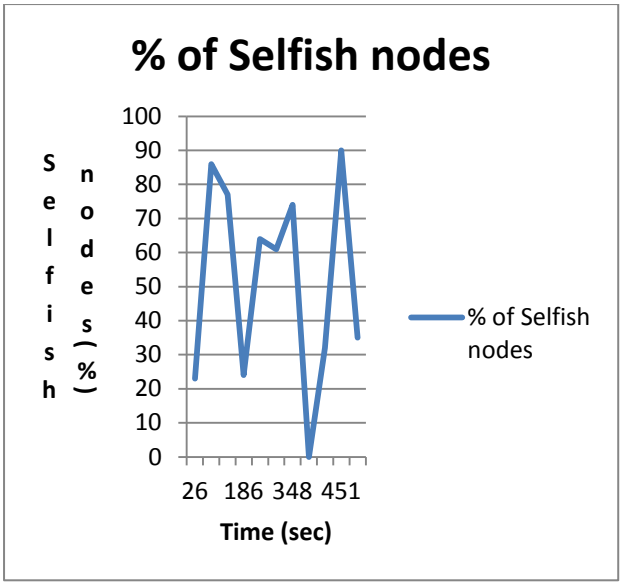

Figure 1: Selfish nodes per Time

Figure shows the % of selfish nodes at certain points of time. The x axis denotes the time taken and the y axis denotes the % of selfish nodes.

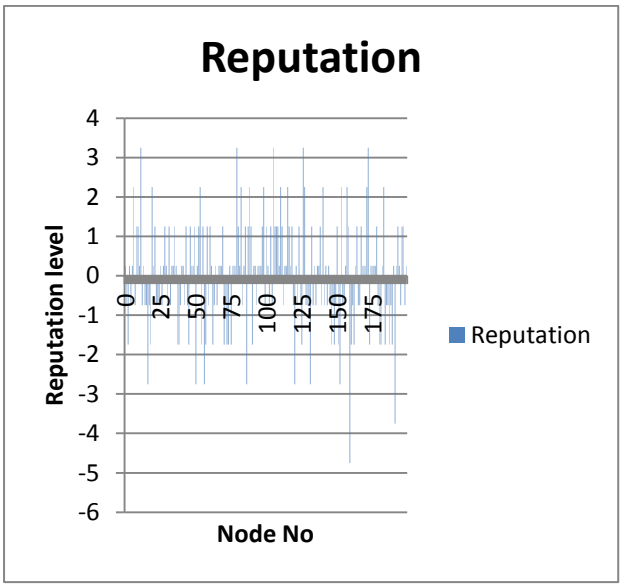

Figure 2: Reputation level of the nodes

Figure represents the reputation level of the nodes with respect to the average reputation level. The x axis denotes the node number and the y axis denotes the reputation level of the corresponding nodes.

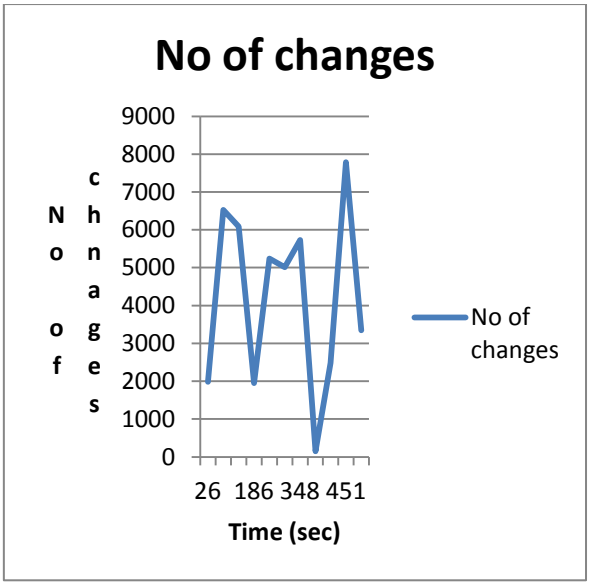

Figure 3 : No of changes in cluster heads

The Figure represents the no of changes performed in the cluster heads. The x axis denotes the time and the y axis denotes the no of changes performed during that time.

4. CONCLUSION

Determination of the cluster head is one of the most important functionalities that has to be performed in an Intrusion Detection system. Finding the appropriate nodes for running the IDS and monitoring the network plays a crucial role in maintaining the consistency of the system. Inappropriate nodes (selfish nodes) have a very high probability of breaking down or providing insufficient process time for the IDS. The current study provides an efficient mechanism for determining the nodes that are to be used for running the IDS. The initial clustering process groups all the nodes in such a way that a single node would be able to maintain all other nodes in its cluster (within proximity). Usage of the Ant Colony Clustering (ACC) algorithm hence provides a heuristic way of determining these clusters, further usage of the Game Theory provides an efficient and continuous mechanism for electing or re-electing the cluster heads.

5. REFERENCES

[1] Lumer E.,Faieta B.,1994,"Diversity and adaptation in populations of clustering ants", in: J.-A.Meyer, S.W. Wilson(Eds.), Proceedings of the Third International Conference on Simulation of Adaptive Behavior: From Animats,Vol.3, MIT Press/Bradford Books, Cambridge, MA, pp.501~508.

[2] Handl, J. and Meyer, B.,2002,"Improved ant-based clustering and sorting in a document retrieval interface",PPSN VII, LNCS 2439.

[3] Maoguo Gong and Liefeng Bo,2007,"Density-Sensitive Evolutionary Clustering", The 11th Pacific- sia Conference on Knowledge Discovery and Data Mining, Springer-Verlag Berlin Heidelberg,pp.507–514.

[4] Ling Chen, Xiao-Hua Xu,Yi-Xin Chen ,26-29 August 2004,"An Adaptive Ant Colony Clustering Algorithm", Proceedings of the Third International Conference on Machine Learning and Cybernetics, Shanghai.

[5] S.Nithya, R.Manavalan ,2012,"An Ant Colony Clustering Algorithm Using Fuzzy Logic", International Journal of Soft Computing And Software Engineering (JSCSE),2012, e-ISSN: 2251-7545, Vol.2,No.5.

[6] Jamaludin Sallim , Rosni Abdullah, Ahamad Tajudin Khader," An Improved Ant Colony Optimization Algorithm for Clustering Proteins in Protein Interaction Network".

[7] O.A. Mohamed Jafar and R. Sivakumar , October 2010,"Ant-based Clustering Algorithms: A Brief Survey", International Journal of Computer Theory and Engineering,Vol. 2, No. 5 , 1793-8201.

[8] Hadi Otrok, Noman Mohammed, Lingyu Wang, Mourad Debbabi, Prabir Bhattacharya ,2008,"A game-theoretic intrusion detection model for mobile ad hoc networks",Computer Communications 31 (2008) 708–721.

[9] Wenying Fenga,Qinglei Zhangc, Gongzhu Hud, Jimmy Xiangji Huange ,2014,"Mining network data for intrusion detection through combining SVMs with ant colony networks", Future Generation Computer Systems.

[10] Chih-Fong Tsai,Yu-Feng Hsu , Chia-Ying Lin , Wei-Yang Lin,2009," Intrusion detection by machine learning: A review", Expert Systems with Applications 36 (2009) 11994–12000.

[11] Shahaboddin Shamshirband , Nor Badrul Anuar , Miss Laiha Mat Kiah , Ahmed Patel ," An appraisal and design of a multi-agent system based cooperative wireless intrusion detection computational intelligence technique", Engineering Applications of Artificial Intelligence 26 (2013) 2105–2127.

[12] D.P.Jeyepalan, E.Kirubakaran ,April 2013,"A Novel Graph Based Clustering Approach for Network Intrusion Detection", International Journal of Computational Intelligence and Information Security, Vol. 4 No. 4,ISSN: 1837-7823.

[13] Qiang, W., Vasileios, M,2005,"A Clustering Algorithm for Intrusion Detection",The SPIE Conference on Data Mining, Intrusion Detection, Information Assurance, and Data Networks Security, Florida, vol. 5812, pp. 31–38.

[14] Joshua Oldmeadow, Siddarth Ravinutala, Christopher Leckie,2004,"Adaptive Clustering for Network Intrusion Detection", Springer-Verlag Berlin Heidelberg,PAKDD 2004, LNAI 3056, pp. 255–259.

[15] XlONG Jiajun, LI Qinghua, TU Jing,2006,"A Heuristic Clustering Algorithm for Intrusion Detection Based on Information Entropy", Wuhan University Journal Of Natural Sciences, Vol. 11 No. 2 2006 355-359.

[16] Maria C.V. Nascimento, Andre C.P.L.F. Carvalho, J,2011,"A Graph Clustering Algorithm Based On A Clustering Coefficient For Weighted Graphs", Brazil Computer Society,17:19–29DOI 10.1007/s13173-010-0027.