Optimization of Inference Intrusion Detection System in Wireless Networks Using Hybrid Genetic Algorithm

Hamed Alavinejad

Department of Information Technology, Islamic Azad University, Ali Abad Katoul Branch, Iran.

Abstract

Nowadays, with the advancement of communication technologies, the establishment of communication needed to set up networks is possible by the aid of different techniques. A wireless sensor network consists of a large number of sensor nodes that have been densely dispersed in the environment. The issue of cooperative intrusion detection in the wireless sensor networks has always been challenging due to the shortage of available energy resources for the participating nodes. Power consumption ratio, network reporting delay time period, network coverage ratio and information quality are the objectives in the wireless network that are contradictory to each other. The issue of determining the type of function and the duties of nodes in intrusion detection has been considered as a multi-objective optimization issue, which aims to minimize power consumption and reporting delay and to maximize the information accuracy and network coverage. By many simulations, scalable solutions are for large networks to solve the problem, and to reach more optimal responses the hybrid genetic algorithm has been used.

Keywords: Wireless Sensor Network, Cooperative Intrusion Detection System, Multi-Objective Issue, Hybrid Genetic Algorithm

Introduction

Wireless sensor networks are a new member of the family of wireless data networks with a series of specific properties and requirements. Nowadays resistance against attacks is one of the challenges of the development of these networks. Regarding the increasing expansion of wireless networks, the establishment of security in these networks is considered an essential point. A wireless sensor network has been composed of a large number of nodes in a specific area, each of which has the ability to collect information such as temperature, pressure, humidity, noise, light, and so on, from the area in which it is located, and sends the collected data to the sink node. Investigating cases such as the implementation of the intrusion detection system in wireless networks, as well as investigating the network objectives such as power consumption, delay ratio in reporting, network coverage, and information accuracy are among the cases addressed in this research.

Problem Statement:

Sensor networks have been developed with the motivation to use in military applications, such as monitoring the battlefield, but today, they are also used in the industry and many non-military purposes. While the presence of wireless sensor networks in military and civilian areas is increasing, the need for security becomes a necessity too. Energy in the sensor network is a vital factor. Reducing energy consumption and increasing the system lifetime in sensor networks is one of the service quality criteria in these systems. So far, many routing algorithms have been presented for sensor networks, and various criteria such as the distance to sink node, traffic load and energy consumption along the route are used to select the appropriate route. Since energy is a vital source, the amount of energy consumption along the route can be a proper criterion for this goal. Security goals for sensor networks include four primary goals of availability, reliability, integrity and applicability like the conventional networks. In this research, the issue of the optimization of the inference intrusion detection system in the wireless networks has been investigated using the meta-innovative algorithms.

Research History

In recent years, security in wireless sensor networks has been considered more. There are various methods to deal with attacks in wireless sensor networks, some of which have been presented in the following. In a research, the security of wireless sensor networks has been investigated. In this article, the goal is to reduce the effect of node delay by combining an efficient competitive model with a WSN cell model. The collaborative model exploits several vulnerabilities existing in the network, such as high node density, putting descending node, and neighbor's intrusion factor to calculate the compromise probability of each cell. Then, it defines the length of its chain for each cell with various timing interval to increase the network's resistance against the node's seizure attack. The proposed plan has been compared with other existing plans, taking into account the

possibility of a key compromise and a number of re-links. In another research to solve the problem of attacks in the wireless sensor networks, a hierarchical framework based on discovery technology and control use is proposed to improve the security of WSNs and at the same time, it is considered with the least complexity and high security needs of WSNs. Continuous decision-making features and dynamic features in UCON can perform continuous attacks using advanced permanent threat detection. In addition, the dynamic compatibility chance discovery mechanism is used to detect unknown attacks. To design and implement a system, an integrated framework has been proposed in which detection of low-level attack is performed with simple rules in the sensor and a high-level attack is performed with complicated rules at the base station. Additionally, networks defined by software and the virtualization technologies of time network function that are identified as low level or high level attacks are used to prevent the attack. A test has been conducted to obtain an attacked data for evaluation. Also, in research, networks have been addressed as infrastructure networks and designing the Hole Defense Systems, that a Genetic Algorithm-Based Implementation and also a PSO method have been proposed, and at the end, a comparison has been performed between GA and PSO. At that time, an intrusion detection system, especially for a black hole attack, is implemented using the Genetic Algorithm / PSO Algorithm. In this article, the charge of battery on each node has been obtained in the first three situations: without DDOS attack, the second: the charge of the distributed battery of each node in the presence of the DDOS attack, and the third: the charge of the distributed battery of each node at the time of preventing the DDOS attack.

Modeling and Work Method

Objective Function

Decision making for the way of distributing intrusion detection system functions in the network leads to the formation of a set of cluster trees that each of them implements a cooperative intrusion detection system.

Separate nodes organized in the cluster trees communicate with each other cooperatively for the intrusion detection purposes and aggregate information. Organized cluster tree affects energy consumption, reporting delay, network coverage, and the quality of data collected. These properties represent every independent optimization goal in the multi-objective issue.

It is assumed that the network of K + 1 node is in the set of $N = \{n_1, n_2, \dots, n_k, b\}$that b is the Base Station and each node in the network has a separate responsibility. Regarding the use of the structure of cooperative intrusion detection system, the nodes are organized as trees of the cluster. All nodes are interconnected with the R radio range. Leader nodes can also communicate with each other with a wider radio range.

$$R' = \beta R \qquad (1)$$

That β is a number greater than one.

A graph composed of leader nodes with the symbol of $G_L$is defined as follows:

$$G_L = \langle L \cup \{b\}, \quad L \times L\rangle (2)$$

In which, if the distance between two leader nodes is less thanR', it means that the two nodes are connected, and if this distance is greater thanR', then the two leader nodes are not connected.

Various cluster tree topologies show different properties of themselves for energy consumption, event reporting delay, network coverage, and information accuracy. Some of these properties, such as energy consumption and delay in reporting should be minimized, while other properties such as network coverage and information accuracy should be maximized. Separately, each of the proposed properties can be considered as a single-objective optimization issue; but the purpose of the cooperative intrusion detection system is to maximize network coverage and information accuracy, and to minimize reporting delay and energy consumption. The objective function of the cooperative intrusion detection system issue is defined as follows:

$$\text{Maximaze}: \left[\left(1 - \frac{P_G}{C_p}\right), \frac{H_G}{C_h}, \left(1 - \frac{D_G}{C_d}\right), \frac{C_G}{C_c}\right] \qquad (3)$$

In which $P_G, H_{G,} D_G,$ and $C_G$are respectively the objective function of power consumption, information accuracy, reporting delay, and network coverage, andC_p, C_h, C_d،andC_c are respectively the optimal values of each single-objective function of the multi-objective issue.

Limitations

With regard to the proposed issue and the structure of the cooperative intrusion detection system, there are limitations governing the issue.

$$\left|L_{T_i}\right| = 1 \quad \forall i = 1, \dots, q \qquad (4)$$

This limitation indicates that only one node leader should exist in any cluster tree.

$$\text{if } n_j \in A_{T_i}, J_{T_i} \text{ then } \text{parent}(n_j) \in A_{T_i} \cup L_{T_i} \forall i = 1, \dots, q \qquad (5)$$

This limitation of the architecture and structure guarantees the cluster tree to match with the structure of the cooperative intrusion detection system.

$$|J_{T_i}| = 1 \quad \forall i = 1, \dots, q \qquad (6)$$

This limitation indicates that at least one connector node must exist in each cluster tree.

$$b_{L_i} \geq b_L^{th} \forall L_i \, (7)$$

This limitation indicates that the remaining charge level in the leader node should be greater than the threshold limit of b_L^{th} itself.

$$T_i \leq |T|_{th} \forall i = 1, \dots, q \qquad (8)$$

This limitation indicates that the size of cluster tree should be smaller than the threshold limit size of each cluster tree in the network.

And finally, the leader nodes must be interconnected in the $G_L$network graph that the $G_L$graph contains the leader nodes and the base station.

Single-Objective of Objective Functions

In order to investigate the solutions of the multi-objective issue of the cooperative intrusion detection system, the constants ofC_p, C_h, C_d،and$C_c$should be obtained from the related single-objective issues, so that the objective function of the main issue is normalized.

Power Consumption

Nodes consume different amounts of energy for communication and computation with respect to the responsibilities assigned to them in the network. Regarding that the aggregator nodes in the aggregate module and the leader nodes in the cooperative module execute complicated functions, they consume more power. Similarly, cluster trees also exhibit different behaviors of energy consumption regarding the number of their nodes and the distribution of nodes type. Therefore, the total power consumption of each cluster tree set may vary from other trees. As a result, the power consumption of all network nodes in the cluster trees in the network should be minimized.

$$\text{Minimize} : P_G = \sum_{i=1}^{k} p_i = \sum_{i=1}^{|L|} p_{L_i} + \sum_{i=1}^{|A|} p_{A_i} + \sum_{i=1}^{|J|} p_{J_i} + \sum_{i=1}^{|O|} p_{O_i} \, (9)$$

Thatp_{L_i}, p_{A_i}, p_{J_i}andp_{O_i}respectively, are power consumption in leader, aggregator, connected, and missing nodes.

$$p_{L_i} = P_{RX_i} + P_{Proc_i} + P_{Rep_i} + P_{Recv_i} (10)$$

The consumed power components in the leader node in the equation 10, are respectively the power consumed to receive the report, the total power consumption to keep the cooperative module active and the variable power for encoding / cooperative information from the child's nodes, the power consumed to transmit the intrusion detection warning, and the power consumed to receive the intrusion detection warning that this equation can be expanded in equation 11 and state its details.

$$p_{L_i} = p_{rx} \sum_{j=1}^{|F_i|} \lambda l_{E_{ij}} + P_{CodC} \left[\lambda l_{E_{ij}} + \sum_{j=1}^{|F_i|} \lambda l_{E_{ij}} \right] + P_{ModC} + \eta l_A (p_{tx} + p'_{tx} + p'_{rx}) (11)$$

The power consumed in aggregator nodes has been presented in the equation 12.

$$p_{A_i} = P_{L_i} + P_{TX_i} = P_{L_i} + p_{tx} \lambda l_{E_i} (12)$$

Unlike the leader and aggregator nodes, the only duty of the connected nodes is to transmit the event vector to the parent and to receive warnings from them. Therefore, the power consumption of the connected nodes is expressed as follows.

$$p_{J_i} = P_{TX_i} + P_{Rep_i} = p_{tx} \lambda l_{E_i} + \eta l_A p'_{tx} (13)$$

Since the consumed power of the missing nodes do not perform any intrusion detection process; so its amount is zero.

Information Accuracy

The way nodes are organized in the cluster tree affects the ratio of information collected. An optimal solution for the multi-objective issue is to collect information as much as possible and reducing the amount of non-useful information collected. It is assumed that for simplicity each node sends only one security parameter as event vector to its parent. The security parameter is a random variable with a Gaussian distribution($x_i \sim N(\mu, \sigma_i^2)$). The amount of information containing x_i is called its entropy, $H(x_i)$ is the data entropy of the reporting parameter of i node. Also, if the i node is the parent of j node, then $H(x_i, x_j)$is the connection entropy of the two variables of $x_i$and x_j.

For cluster trees, the information function of maximizing the total entropy data of network nodes has been introduced.

$$\text{maximize}: H_G = \sum_{i=1}^{k} h_i \qquad (14)$$

That $h_i$is the data entropy of connected i node, and $h_i = H(x_i, \dots, x_n)$is the entropy of all the parameters reached by the aggregator i along with its security parameter. Each observed parameter includes$(x) = \frac{1}{2} * \log(2\pi e\sigma^2)$ of the bit of information, and $\sigma^2$is the variance of the random variable of x. The n entropy of the Gaussian random variable is:

$$H(x_i, \dots, x_n) = \frac{1}{2} * \log[(2\pi e)^n \det(\Sigma)] \quad (15)$$

That the Σ is the covariance matrix of connected variables.

$$\sum\nolimits_{x_1, x_2, \dots, x_n} = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \cdots & \rho_{2n}\sigma_2\sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1}\sigma_n\sigma_1 & \rho_{n2}\sigma_n\sigma_2 & \cdots & \sigma_n^2 \end{pmatrix} \qquad (16)$$

That $\rho_{ij}$is a function of the place of i node and j node. Regarding that the attacker's place in the network has been considered as standard and independent distribution from the node place, so there is the equation of 17 for the coefficient of ρ_{ij}.

$$\rho_{ij} = \begin{cases} (A_i \cap A_j)/A & \text{if} \quad d_{ij} \le 2R \\ 0 & \text{otherwise} \end{cases} \qquad (17)$$

The overlapping area of these two nodes is calculated as follows.

$$A_i \cap A_j = 2R^2 \cos^{-1}\left(\frac{d_{ij}}{2R}\right) - \left(\frac{d}{2}\right)\sqrt{4R^2 - d_{ij}^2} \qquad (18)$$

Reporting Delay

In the cooperative intrusion detection system model, it is assumed that the aggregator node of all received reports is processed in the time range of τand sends an aggregated report to its parent in the time range of $\tau + 1$. The delay between the reports sent by a node up to the time when it is located by a leader node in the hub μ_i interval is equal to the number of hubs. Therefore, the event reporting delay in the network is defined as follows.

$$\text{Minimize}: D_G = \sum_{i=1}^{k} \mu_i \qquad (19)$$

Network Coverage

The purpose of this research is to maximize network coverage. A node that is the member of the cluster tree of T_i, regardless of its duty, has been covered in the network. Therefore, only the missing nodes are not covered in the network. So, the objective function of the node coverage in the network is defined as follows.

$$C_G = \sum_{i=1}^{q} \sum_{j=1}^{|T_i|} c_{ij} = k - |O| \quad (20)$$

Describing the Conditions of Proposed Algorithm

The optimization issue described in the previous chapter has the properties that limit its solving methods. The number of possible cooperative topologies grows significantly with the growth of the number of nodes. Also, the issue has non-linear constraints and the objective function is of discrete functions. To cope with the challenges proposed by evolutionary algorithms to solve the issue, optimization has been used. Intrusion detection systems are the basic part of the wireless networks of limited resources. Due to the shortage of focal points that network traffic is analyzed, wireless communication channels provide limited bandwidth. Most researchers have proposed the dispersed structure for implementing an intrusion detection system that this architecture is implemented by placing an intrusion detection agent in each node.

Intrusion Detection System Structure

The proposed issue is the intrusion detection in the wireless networks that the target system consists of the wireless networks of limited resources along with a separate base station. The network has been synchronized irregularly in respect of time and all nodes know their places. The base station has the capability of heavy computations and regularly collects network information and uses it to implement the functions required to execute intrusion detection functions. This station also decides on the execution place and time of the intrusion detection function. This decision making is performed regularly or when the remaining network energy reaches the threshold limit and changes the network conditions.

The network is always illustrated as tree or cluster trees. Regarding the introduced structure, each one of the nodes of the cluster tree has duties.

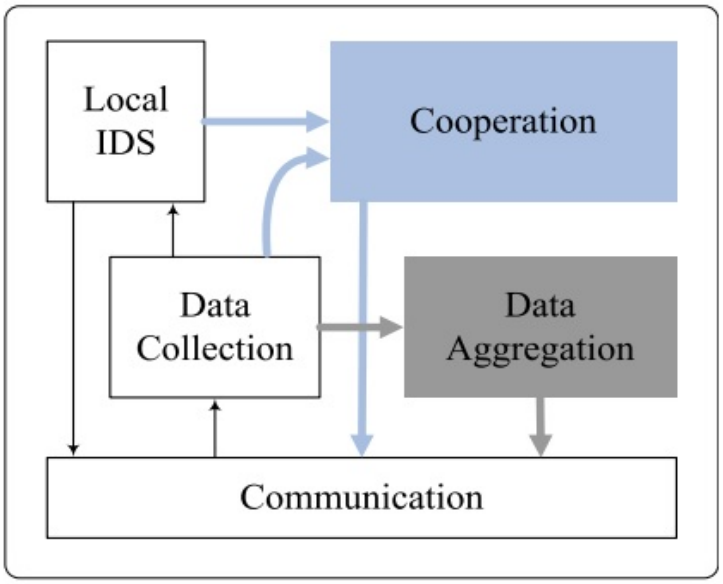

Figure 1: General Structure of Intrusion Detection System

Hybrid Genetic Algorithm

A hybrid genetic algorithm is a compromise between the optimality of solutions and the speed of obtaining solutions. In large networks, this algorithm is capable of obtaining better responses than a genetic algorithm over a shorter period of time. The hybrid genetic algorithm runs in two phases to reach the optimal response. In the first phase, named as r-hop selection leader node, a set of nodes is selected as the leader node that connects to a number of followers and has more charge level. Then, in the second phase, the optimal assignment of the role is executed for the follower nodes of each leader node. In Figure 2, a hybrid genetic algorithm pseudo-code has been presented.

1. Find F_i for $\forall_i$
2. $L = \{\}$
3. While $o \neq \emptyset$ do
4. $L_i = \max_{i \in N} \{|F_i| \times b_i\}$
5. $L = L \cup \{L_i\}$

6. $O = O - F_{L_i}$
7. update − flowers
8. End while
9. Modify – Clusters()
10. For i=1 to |L| do
11. Run GA on $L_i \cup F_i$
12. End for

Figure 2: Hybrid Genetic Algorithm Pseudo-Code

The first phase in this algorithm is in a way that at first the amount of followers of each node is computed. Then, in a repetitive process, based on the level of battery charge and the number of followers, the leader nodes are selected. Then the selected node is added to the leader nodes set and all its neighbors are removed from the set of missing nodes. Since the addition or deletion of the nodes causes a change in the number of neighbors for the remaining nodes, the algorithm calculates the amount of followers of each unselected node again. The algorithm performs the selection of the leader until no missing node remains. It is worth mentioning that according to the network conditions, it is possible that a single-node cluster is created, which means a missing node, that in such conditions, the clusters are corrected, and the missing node will attach to the smallest cluster. So the network coverage is guaranteed by clusters.

In the second phase of the algorithm, the clusters are considered as small separated networks, and for their nodes to assign their duties in the cooperative intrusion detection system, the genetic algorithm is executed for each cluster.

Results and Findings

The hybrid genetic algorithm may have more optimal responses than the genetic algorithm in the large-sized networks. In Figures 3 and 4, respectively, the cluster trees generated from the single-objective information issue in a random 49-node network and 49-node grid network have been illustrated using a hybrid genetic algorithm.

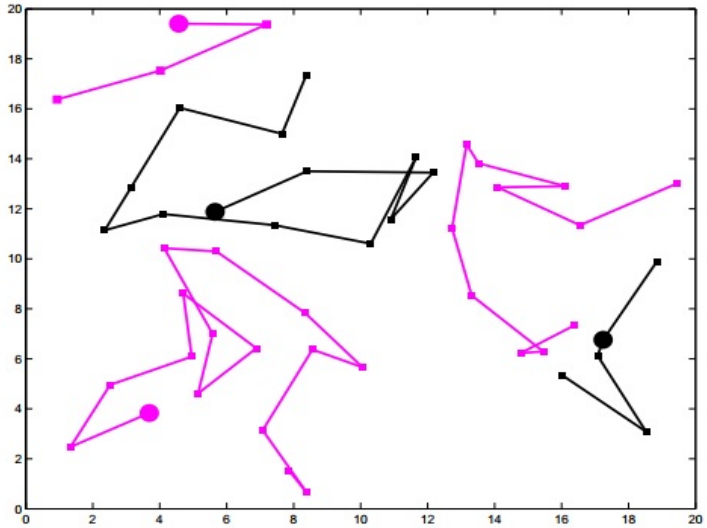

Figure 3: Cluster Trees of Single-Objective Information Issue in the Random 49-Node Network of Hybrid Genetic Algorithms

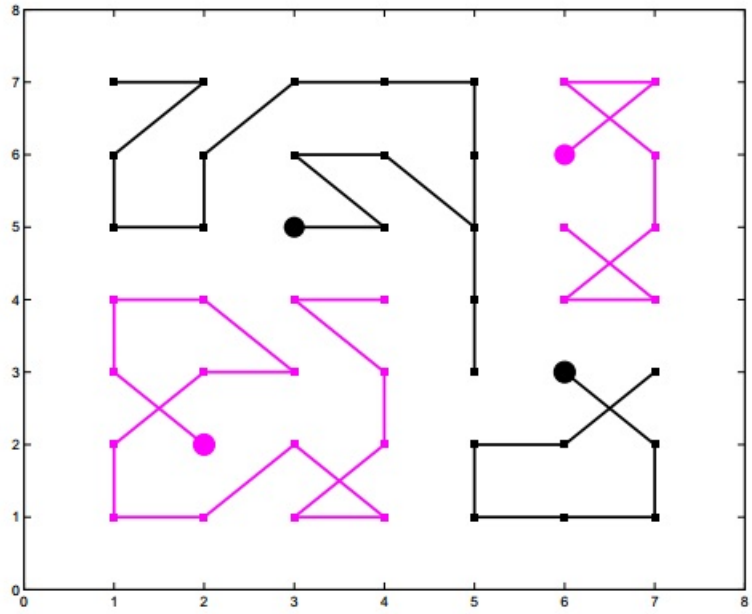

Figure 4: Cluster Trees of Single-Objective Information Issue in the 49-Node Grid Network of Hybrid Genetic Algorithms.

From cluster trees generated in Figures 3 and 4, it can be understood that they have similar characteristics, and have long linear topologies with many aggregator nodes. The optimal response of the hybrid genetic algorithm does not have clusters with one or two nodes. In Figures 5 and 6, the cluster trees generated from the penalized multi-objective issue of random 49-node network and 49-node grid network have been displayed, using the hybrid genetic algorithm.

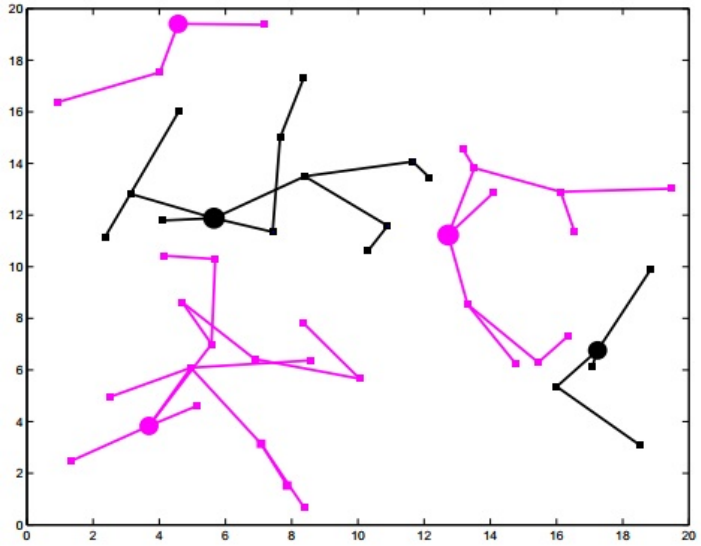

Figure 5: Multi-Objective Cluster Trees in the Random 49-Node Network of Hybrid Genetic Algorithm

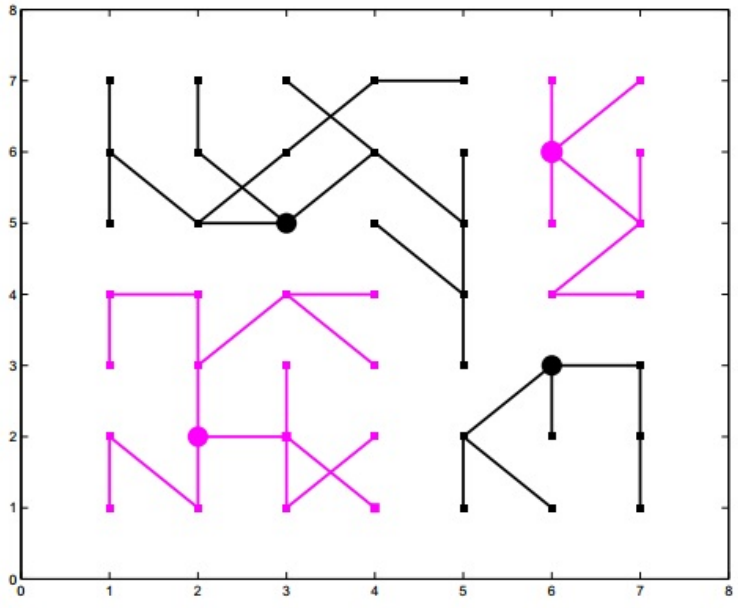

Figure 6: Cluster Trees of Multi-Objective Issue in a 49-Node Grid Network of Hybrid Genetic Algorithm

From the cluster trees generated in figures 5 and 6, it can be understood that the cluster trees are completely separated and there is no overlap and linear intersection between clusters. It can be also concluded that the hybrid genetic algorithm has a stronger control over the resistance of cluster trees.

Performance Comparison of Genetic Algorithm with Hybrid Genetic Algorithm

According to the previous findings [4] and considering that the genetic algorithm has a high connection with the number of generation and population size, for having a more complete comparison, the size of the initial population has been considered twelve times of the network size and the repetition number of the generation has been regarded $\min\{6000, 2 \times k\}$. In Figures 7 and 8, the issue of execution time for random networks has been respectively shown.

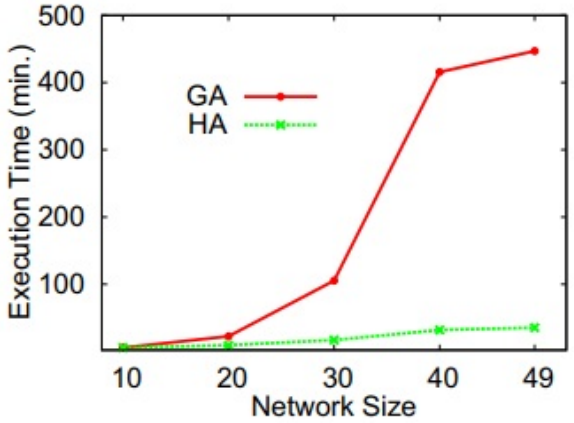

Figure 7: Comparison of the Execution Time of Proposed Algorithms in Random Networks

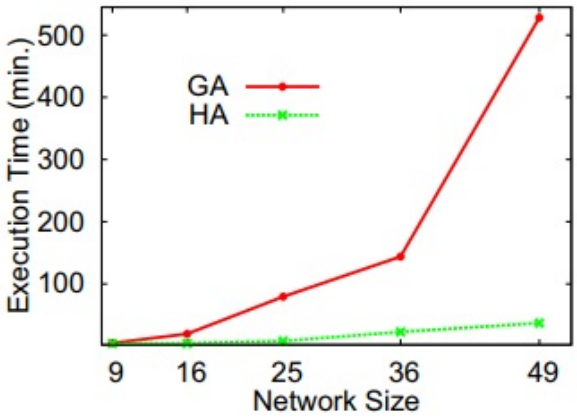

Figure 8: Comparison of the Execution Time of Proposed Algorithms in the Grid Networks

Also, in Figures 9 and 10, the value of the objective function of information has respectively been displayed in the form of a single-objective issue for various networks.

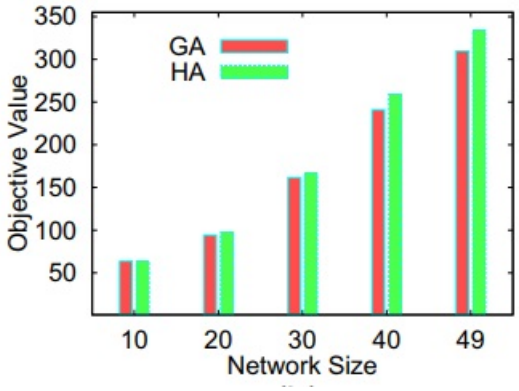

Figure 9: Comparison of the Value of Objective Function of Proposed Algorithms in the Random Networks

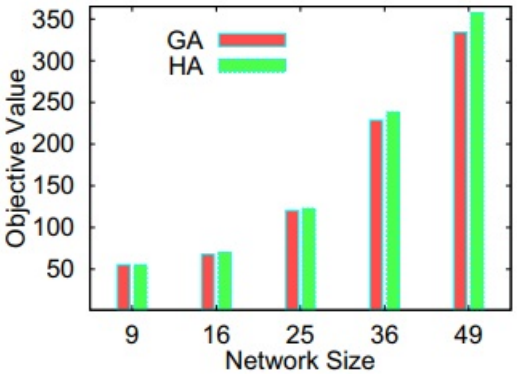

Figure 10: Comparison of the Value of Objective Function of Proposed Algorithms in the Grid Networks

As it can be understood, the execution time and convergence of the hybrid genetic algorithm is much lower than the genetic algorithm, and to compare the value of the objective function in the small-sized networks, there is not that much difference in the final result, while in large networks, the hybrid genetic algorithm has found a more optimal response.

In the multi-objective issue and considering the objective function of power and delay as a constraint, due to the high dependence between power, delay and information, the hybrid genetic algorithm cannot generate optimal responses.

The results obtained from the simulation of various networks that have been proposed and investigated indicate that the use of hybrid genetic algorithm improves the convergence time and the value of the objective function in larger networks.

References

[1] Mosaddegh Moghaddam, Shobeir (2014), "Airport Navigation Visual Assistance Systems", Tehran: Publication of Bisheh
[2] MohammadiMoteh, Bahram (2000), "Visual Navigation Assistance Devices (Bachelor Thesis)", Faculty of Civil Aviation Industry, Tehran
[3] Alamdari; Ali Akbar; Alamdari, Nasrin; "The Most Complete Educational Reference of MATLAB" Tehran: Publication of NegarandehDanesh
[4] Hamed Alavinejad "Optimization of Inference Intrusion Detection System in Wireless Networks Using Genetic Algorithm", Islamic Azad University, Ali Abad Katoul Branch
[5] Annex 14, Aerodromes,Valume I. Aerodrome Design and Operation .Fifth Edition, July 2009.
[6] Federal Aviation Administration, AIRPORT LIGHTING EQUIPMENT CERTIFICATION PROGRAM, 2012.
[7] Shyama Prosad Chowdhury, Karen Rafferty, and Amit Kumar Das., (2010) Localisation and Tracking of an Airport's Approach Lighting System. (Doctoral dissertation), L. Bolc et al. (Eds.): ICCVG 2010, Part I, LNCS 6374, pp. 19–26, 2010.Springer-Verlag Berlin Heidelberg 2010.
[8] Shyama Prosad Chowdhury, Karen Rafferty and Stuart Ferguson, (2010), Simulation and Performance Assessment of Airport Landing Lighting (Doctoral dissertation), 2010 13th International IEEE Annual Conference on Intelligent Transportation Systems Madeira Island, Portugal, September 19-22, 2010