

A Genetic Algorithm-based Optimized Fuzzy Adaptive Path Selection in Wireless Sensor Networks

Muhammad Akram¹, Muhammad Ashraf¹, Tae Ho Cho²

Abstract:

In Wireless sensor networks, energy efficiency can be achieved by adaptive choice of the data forwarding path to balance the energy dissipation in the network. This adaptive path selection is done through a fuzzy rule-based method given the input parameters. Due to uncertainty in reasoning and inferencing process and imprecision in the data, the fuzzy-based system becomes an ideal choice for the selection of the paths. In fuzzy systems, the membership functions need to be optimized to make the best use of the fuzzy inferencing and improve the performance of the fuzzy system. Genetic algorithm-based fuzzy membership function optimization technique selects the optimal solution in a feasible time and saves from the hassle of manual intervention. Manual optimization efforts are unfeasible for common applications and take unlimited time and human expertise to optimize functions in an exhaustive search field. This technique assesses the fitness of the membership functions through simulation outcomes and optimizes them through genetic algorithm based evaluation process. The proposed scheme consists of three modules; The first module simulates the membership function in the given network model, the second module analyzes the performance efficiency of the membership functions through simulation, and the last module constructs the subsequent membership-function populations using GA techniques. The proposed method automatically optimizes the membership functions in the fuzzy system with little human intervention, requires minimal human expertise and saves ample time in the optimization process.

Keywords: *Fuzzy optimization; Route selection; Filtering; Genetic algorithm; Fuzzy.*

1. Introduction

Wireless sensor networks are economically a feasible tool for monitoring physical world, usually comprising of a many number of sensor nodes, and they are often densely installed in hostile locations [1]-[4]. Sensor networks are expected to function as a flexible, universal and effortlessly installable solution for cyber-physical systems and applications [5]. Sensor nodes are usually placed unattended in exposed surroundings, therefore, they are extremely susceptible to being physically captured and compromised [6]. Nodes in the network can be compromised and exploited by the attacker to insert false sensing

information into the network, which both causes false alarms at the base station (BS) and waste the scarce energy resources of the data forwarding nodes in the network [7]. Similarly, compromised nodes are misused by the attackers to insert counterfeit message authentication codes (MACs), alternatively referred to as votes, to cause the dropping of the legitimate reports at the intermediate nodes during the en-route verification process [4], [7]. Injection of false votes restricts the true information from reaching the BS. As shown in Fig. 1, an attacker can exploit a compromised node to either inject fabricated data with false votes or attach false votes to legitimate data to

¹ College of Information and Communication Engineering, Sungkyunkwan University, Suwon 16419, Republic of Korea.

² College of Software, Sungkyunkwan University, Suwon 16419, Republic of Korea
Corresponding Email: akram.khan@skku.edu

cause the aforementioned threats to the network. Researchers have suggested various security schemes to filter fabricated data en-route, and deliver the legitimate information to the BS given the number of fake votes attached to a legitimate report is less than a certain value [[3], [4], [7]-[19].

PVFS and FASIN filter fabricated reports en-route before they consume significant energy resource of the en-route nodes [4], [7].

PVFS and FASIN also allow true reports with false MACs, which are less than certain number, to be delivered to the BS. In PVFS, every cluster member node share their own authentication keys with the verification nodes probabilistically whereas FASIN employ a fuzzy inferencing system installed at each CH which helps to adaptive select the verification on the data forwarding paths [3], [7]. In minimum cost forwarding-based (MCF) schemes [[10], routing paths are chosen by only considering the distance and the energy efficiency of the routing paths. Constant uses of shortest or minimum cost paths leads to uneven work load across the network and causes network partitioning [11]. In [11], we recommended fuzzy based adaptive selection of data routing path to facilitate energy balancing across the network and avoid network partitioning. Fuzzy adaptive choice of data routing paths feasibly balance the work burden between different available routes. Load balancing spreads the utilization of the energy resources in the network, potentially lending to an extended network lifetime [12]. The verification nodes on all the available paths are equal in number which implies that the filtering capability of all the routing paths is essentially same. The adaptive filtering route selection achieves extra energy saving and results in network lifetime extension. In fuzzy adaptive path selection (shortly referred to as FAPS hereafter) [11], a fuzzy rule-based inferencing system chooses among the available data forwarding paths, considering the input parameters such as the verification nodes' average distance, the distance of the

routing path, and the average residual energy resources on the data routing path. To exploit FAPS in a real-world WSN, the fuzzy membership functions in the fuzzy inferencing configuration are required to be optimally tuned for the actual network configuration. However, such an optimization is usually beyond human skills [13]. Even if it is likely to manually optimize the system, the time-cost of manual optimization would be unpractical for real-world WSN configurations.

With prior knowledge it is easy to understand and construct the IF-THEN structure of fuzzy rules. However, many parameters are specified by experts and the identification of these parameters can be treated as optimization problem. Several optimization techniques have been proposed. However, all these techniques do not always guarantee the finding of an optimal solution in multi-parametric space [13]. A genetic algorithm based membership function optimization technique has proved to be more reliable in discovering the ideal solution in multi-parametric space [13]. A genetic algorithm (GA) is a concurrent, global search method that imitates natural genetic processes [14]. GA becomes an ideal choice for membership function optimization due to its ability to simultaneously explore multiple points in the search space and it converge to a near optimal solution. The search space does not need to be necessarily differentiable and continuous for the GA to work, and it has the capability to iterate several times on each piece of received data [13].

2.Fuzzy Adaptive Path Selection (FAPS)

In [11] we presented a fuzzy rule based adaptive path selection scheme (FAPS) which also makes use of fuzzy adaptive selection of verification nodes (FASIN) [4]. In FAPS, each intermediate path between the source cluster and the base station (BS) have equal number of verification node which are initially selected probabilistically. FASIN [4] use fuzzy rule-based inferencing to select verification nodes

on each path and gradually improves their average distance to the source cluster. FAPS selects the suitable routing path considering the following input parameters for each path.

1. The number of verification nodes with in L hops on the path ($L = \text{cluster size}$)
2. The path distance
3. The average residual energy of the nodes on the path.

In FAPS, the membership functions of the input and output parameters are carefully selected to obtain desirable results based on the simulation outputs. These membership functions can be modified and adjusted based on the professional understanding of the network parameters such as network size and

density of nodes. However, in real-world WSNs, manual optimization of the input membership functions requires a lot of efforts and expert knowledge and is mostly impractical as it may not obtain an optimal solution within reasonable time. Such an optimization task is usually beyond human expertise and capabilities because the space has to be exhaustively searched for the optimal solution.

We propose to optimize membership functions for FAPS in WSNs using genetic algorithms. The primary goal of GA-based fuzzy adaptive path selection (GAFPS) is to provide fuzzy membership functions that are optimized and function well in accordance with the network configurations requiring little human expertise and involvement.

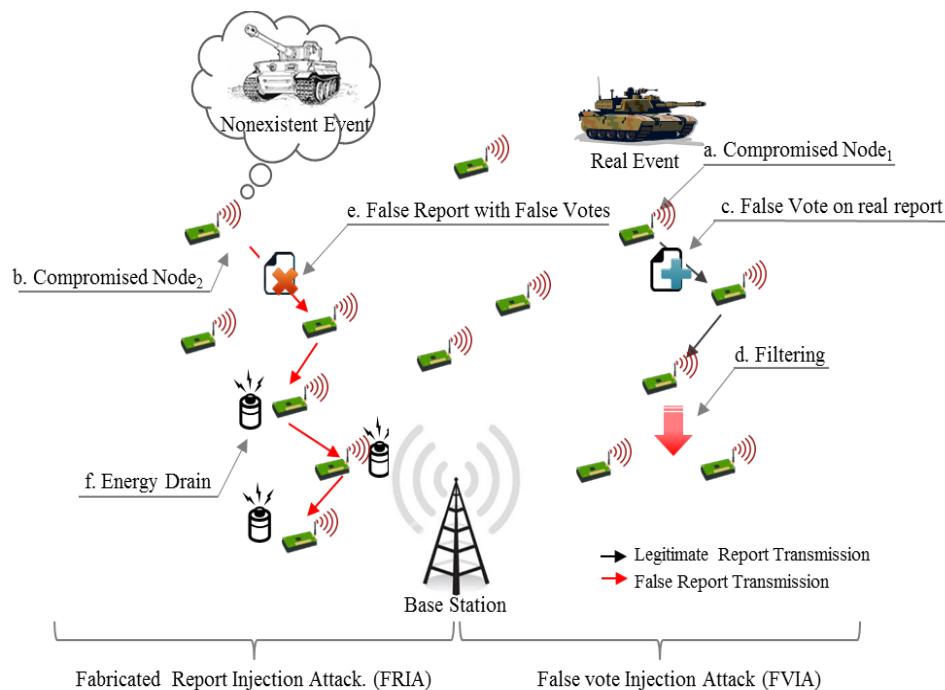


Fig. 1. False report and false vote injection attacks in wireless sensor network.

3.GA-based Fuzzy adaptive path selection (GAFPS)

This section describes the Genetic algorithm-based optimization of the fuzzy input parameters for the fuzzy adaptive selection of data routing paths.

3.1. Genetic Representation of Membership Functions

In GA-based membership function optimization scheme, a single chromosome is used to represent the set of fuzzy input membership functions during a single trial.

The input parameters of the fuzzy rule based system in [11] are:

1. REL
2. NKİH
3. DIST

The output parameter of the fuzzy rule based system is:

1. Fitness

Each input parameter has 3 triangular membership functions which are labeled as follows:

1. REL: L (Low) , M (Medium), H (High)
2. NKİH: Le (Less), Mo (Moderate), Mr (More)
3. DIST: N (Near), F (Fair), Fa (Far)

Therefore the fuzzy rule-based system makes use of 27 ($3 \times 3 \times 3$) fuzzy rules. Similarly, the output of the fuzzy rule-based system produces the inferred fitness value of the path for which the fuzzy inferencing is being carried out. The output parameter consists of 4 triangular membership function denoted by:

1. Fitness: P (Poor), Mo (Moderate), B (Better), Be (Best)

The design of the membership functions should be such that they satisfy two conditions given below:

- a. Every membership function intersects only with the nearest adjacent membership functions;
- b. Membership values should be equal to or nearly 1 in related fuzzy sets.

Membership function can be conveniently represented through chromosomes. Chromosome are binary coded comprising of bits: 1s and 0s. Each bit in a chromosome represents a gene value which is either 0 or 1. A single chromosome represents a single trial set of input fuzzy membership functions as depicted in Fig. 2. One parameter suffices to represent 3 membership functions of triangular shape for a single input. Therefore, 3 parameters for the input membership functions are coded into a chromosome for each input variable for its representation in a chromosome. Each parameter of REL, DIST are coded by 6 binary digits whereas the single parameter of variable NKİH is coded by 3 binary digits. Therefore, a single chromosome size is $(6+6+3) = 15$ bits.

The optimization process is carried out in design-time prior to network deployment. Exhaustive search strategy may not find an optimum solution within bounded time; it takes $2^{15} (=32768)$ trail sets and each set is simulated for its performance evaluation. We don't modify the fuzzy rules and they are fixed. Modifying the fuzzy rules also widens the search space which increases the time cost exponentially, and we cannot expect a near optimal solution [15]. In GAFPS, 300 generations, each comprising a population of 30 chromosomes, are created. Therefore, only 9,000 trial sets are tested in simulations. Fig. 2 shows the representation of the input fuzzy membership function through a single chromosome.

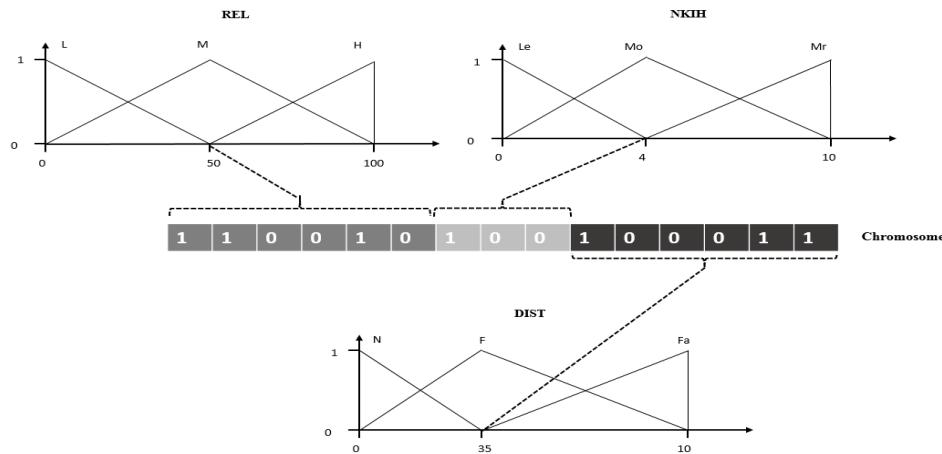


Fig. 2. Genetic representation of membership functions.

3.2. GA Module

The GA module stores a collection of chromosomes that are represented by a sequence of binary strings. Initially, the bit values in a single chromosome are set randomly with a probability value of 0.5 for setting each bit in the chromosome. Each chromosome is tested through simulations for their fitness values. Subsequent generations of chromosome are chosen according to evolutionary techniques employed in genetic algorithms such as selection, crossover, mutation etc. Fig. 3 show the evolution process and explains the construction of the subsequent generations of chromosomes.

A random pair of chromosomes is chosen for reproduction in the next generation according to their probabilities. Genetic operator such as selection, crossover and mutation help to boost chromosome up towards the optimum. Selection realizes the survival of the fittest in the evolution process. Fig. 3 shows the evolution of a 4-parameter (for example) input fuzzy membership function. Fig. 3(a) illustrates the selection of a chromosome in the i^{th} generation to be reproduced in the $(i+1)^{\text{th}}$ generation. GAFPS uses the fitness proportional method for the selection of a pair of chromosomes to reproduce their next generation offspring. In fitness proportional method, the selection

probability of a chromosome for reproduction is directly proportional to its fitness value. The chromosome, whose fitness value is greater, is highly likely to be selected, therefore the average fitness value of the entire population improves over generations. A chromosome with the higher fitness value can be selected more than once whereas the chromosome with the least fitness value may not be chosen at all. In Fig. 3(a), chromosome c4 has the highest fitness value and has been selected twice whereas chromosome c2, with the lowest fitness value, has not been selected at all. Genetic operator, cross over, imitates the genetic inheritance in the offspring and recombines segments of the individuals corresponding to parents. It ensures the exploration of search space.

Fig. 3(b) depicts the crossover operation. Crossover produces a swap or a shift of the fractional parameters to produce new membership functions as depicted in Fig. 3(b) and Fig. 3(c), respectively. Fig. 3(b) shows the interchange of a single parameter between $c1'$ and $c2'$, $c2'$ and $c3'$, and $c3'$ and $c4'$ whereas Fig. 3(c) shows the shift in the membership function achieved through partial altering of the membership function. The altering of the parameter can also be done through mutation which imitates the genetic mutation (self-variation) of genes. The mutation operator is

another means of exploring search space and only a small number of genes (bits) must be altered in a chromosome in accordance with a designed probability. Fig. 3 (d) shows the mutation operation wherein few bits of the chromosome c4' are partially altered to produce a new membership function.

Optimization process can be done more than once to avoid convergence to local optima/maxima. Therefore, we have carried out the optimization process twice changing the crossover and mutation probability values. In the first round of optimization, probability values of 0.9 for crossover and 0.01 for mutation are used. In the second round of optimization, the crossover and mutation probability values are changed to 0.6 and 0.1 respectively. GAFPS selects the chromosome that has a greater fitness value.

3.3. Simulation Module

During each generation, every chromosome is evaluated for its fitness value through a simulation run in the simulation module. Simulation module simulates the target WSN for each single chromosome. The Simulation module comprises of two models i.e. a network model and an attack model. Fig. 4 depicts the simulation process in the simulation module. When a chromosome arrives at the simulation module, it breaks it into 3 segments and reconstructs the fuzzy sets. The fuzzy membership functions are supplied to the rule based system of the FAPS as in [11] in the simulation module. The fitness evaluation module uses results of each simulation run to calculate the individual chromosome's fitness value. During each simulation run, we measure the residual energy of the network, the average number of hop traversed by a report before a first false MAC is detected, and the average number of hops traversed by false reports before they are filtered.

3.4. Fitness Evaluation Module

Fitness evaluation module calculates the fitness of every individual chromosome using the results produced by the simulation unit. A chromosome c_{ij} 's fitness value $F_{c_{ij}}$ is calculated by the following equation:

$$F_{c_{ij}} = (\alpha \times (1/n \sum_{i=1}^n REL_i)) + (\beta \times \frac{1}{H}) \quad (1)$$

In the above equation, c_{ij} is the i^{th} chromosome in the j^{th} generation, REL = the residual energy of the network, n = size of the network, H = the average number of hops traversed by false reports, and α and β are weight factors.

The fitness value calculated by the fitness evaluation module is provided to the GA module. The value of the weight factors are determined with regards to the objective of the target WSN. Since, the energy of the network is the most important factor, therefore value of α is usually much greater than the values of β [16]. Since the detection power of all the paths is same in terms of the number of verification nodes, the drop rate of false reports and delivery success rate of legitimate reports are irrelevant and not considered in calculating the fitness value of a chromosome. We are only interested in improving the energy saving of the sensor network and early filtering of the false reports.

3.5. Optimized Fuzzy Adaptive Path Selection

The GA module contains a population of 30 chromosomes in each generation. The simulation runs and regenerates a new population using genetic operations until the terminal condition reaches i.e. 300 generations. Total 400 reports are generated during each simulation run including legitimate reports with true MACs, legitimate reports with false MACs, and false reports. Fig. 5 illustrates the optimized input fuzzy membership functions.

A randomly deployed static network with 4000 homogenous nodes is simulated in a custom simulator developed in Microsoft C++ 2012. The network is contained in a $1200 \times 1000 \text{ m}^2$ 2-D terrain. The cluster size is $L=10$, and every node in the cluster possesses a single key for authenticating the report. If the threshold value T_f reaches 3, the report will immediately be dropped, whereas a report is accepted as a legitimate report if the value of T_f reaches 4. Each report is authenticated by 5 nodes randomly picked up in the cluster

including CH. 16.25/12.5 μJ of energy per byte are consumed to transmit/receive, and 15 μJ are consumed to generate a vote. The size of the report and MAC is 36 and 4 bytes, respectively. MAC generation consumes 15 μJ of energy. The verification of a report at a

particular node consumes 75 μJ of energy. The BS is located at the upper left edge of the network. The network is exposed only to FVIA and FRIA attacks. Cluster based organization restricts compromised nodes in different clusters from colluding with one another.

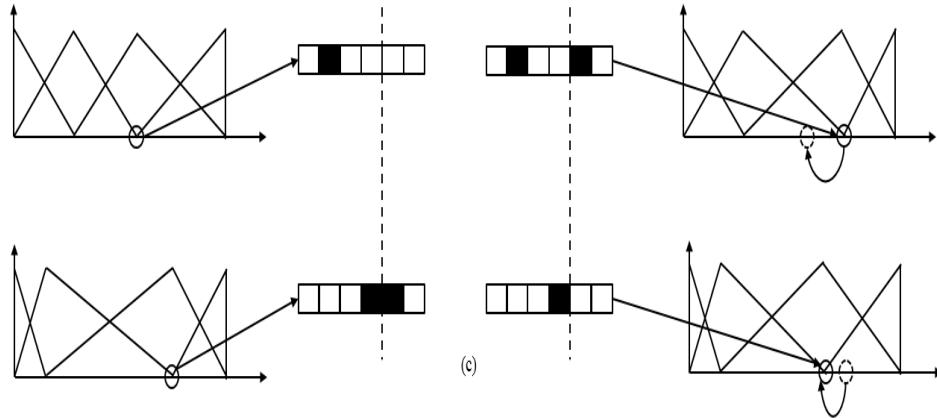
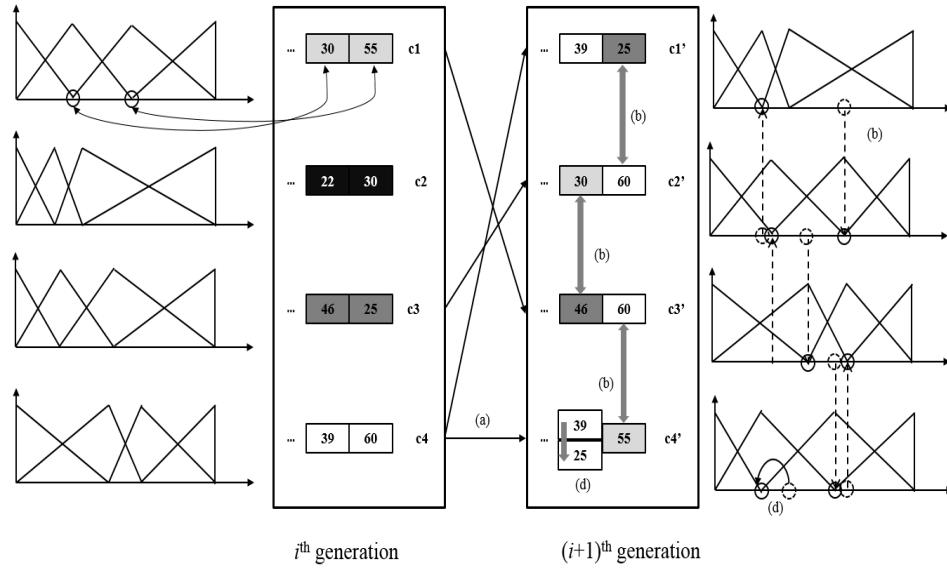


Fig. 3. Evaluation of next generation of chromosome population.

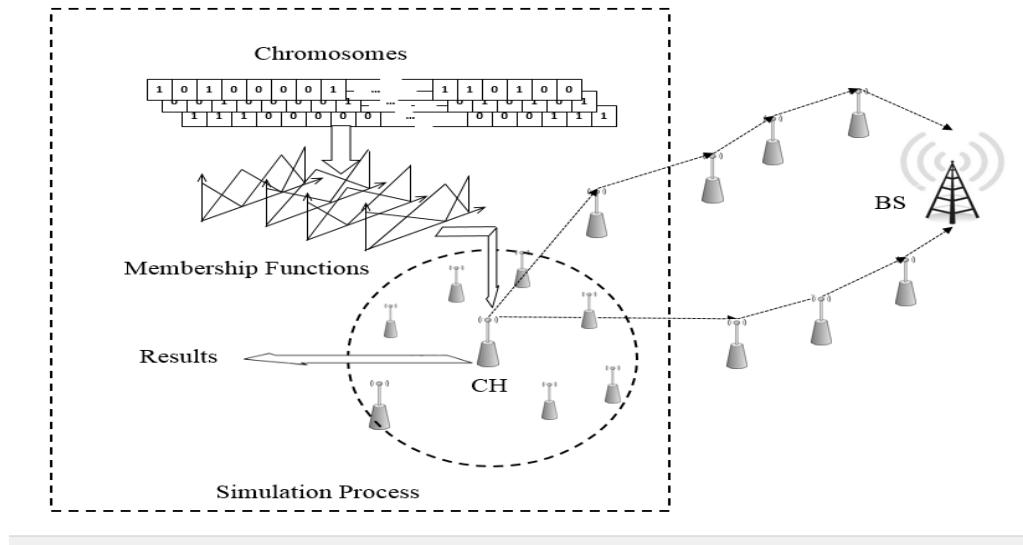


Fig. 4. Simulation Process.

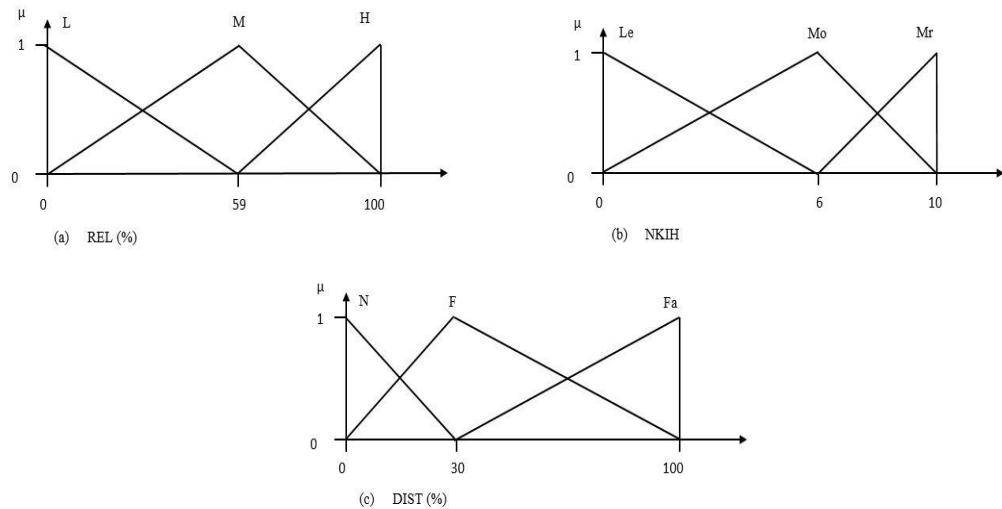


Fig. 5. Optimized input fuzzy membership functions.

4. Simulation Results

Fig. 6 shows the average energy invested on a false report being forwarded to the BS before getting filtered during different generations. As illustrated in Fig. 6, the average energy consumption per filtered report in 1st and 40th generations is greater than in the Final generation. The reason for higher energy consumption in 1st and 40th generation is that the membership functions are chosen such that it may choose a path with farther verification nodes than a path which has more nodes closer to the source cluster.

An arbitrary and haphazard design of membership functions chooses a less suitable path, and hence an increase in energy dissipation, and irregular pattern in the energy curve. This analysis reveals that the careful selection of membership functions significantly impacts the performance of the network. Since the membership function are optimized in the final generation therefore, it exhibits a steady behavior and the energy consumption is lesser than in non-optimized generations. The energy consumption of the final generation compares well with that of the

manually optimized solution and in fact save further energy.

Fig. 7 shows the improvement in the average fitness value of the chromosomes in the population during each generation. The values of the chromosomes are generated randomly during the 1st generation and GA technique such as crossover, mutation and selection are applied in each generation to improve the average fitness of the population. As shown, the average fitness value of the population improves and eventually stabilizes.

Fig. 8 shows that the optimization of fuzzy membership functions also helps to improve the performance of the proposed scheme with regard to the average number of hops traversed by a report before the first false MAC is detected. Optimized membership functions ensure the selection of the appropriate and most feasible route for data forwarding. Therefore, the energy saving capability of the network also improves and false reports are filtered relatively earlier.

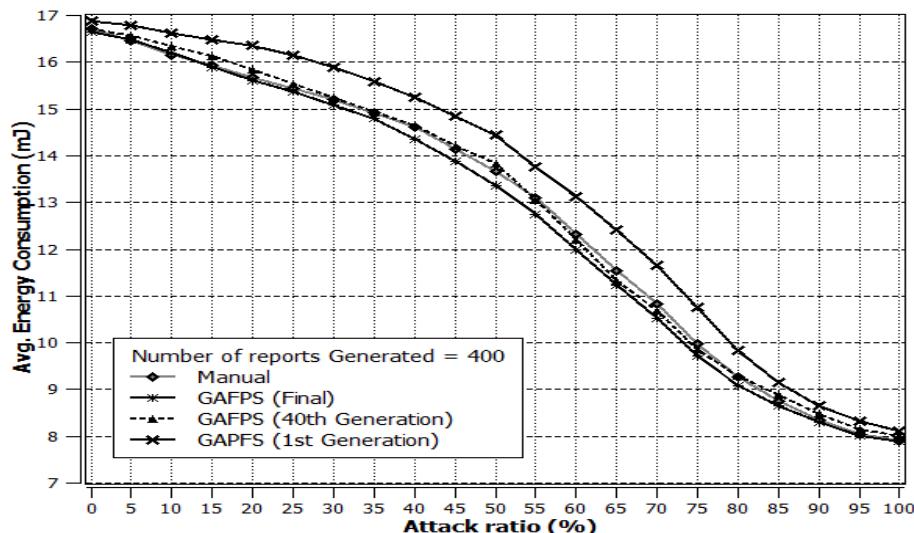


Fig. 6. Average Energy consumption per report before getting filtered.

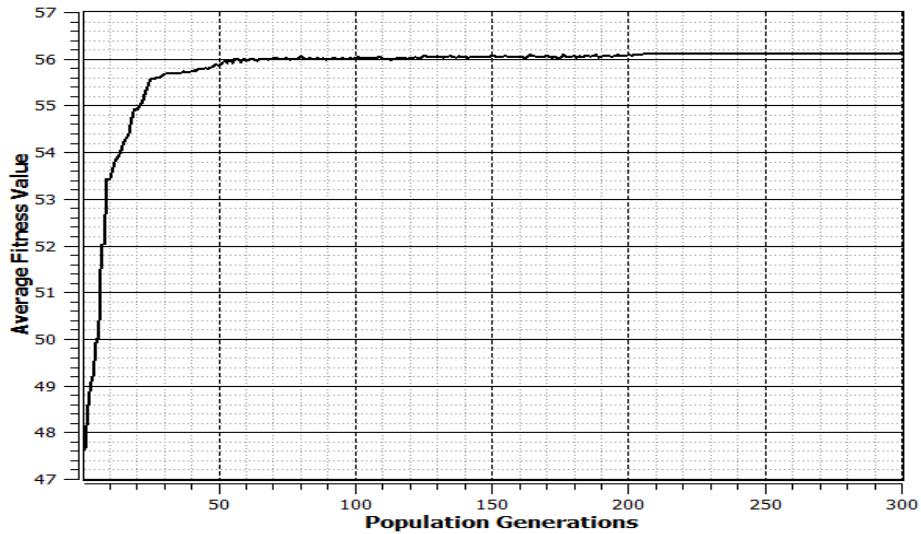


Fig. 7. Average fitness value.

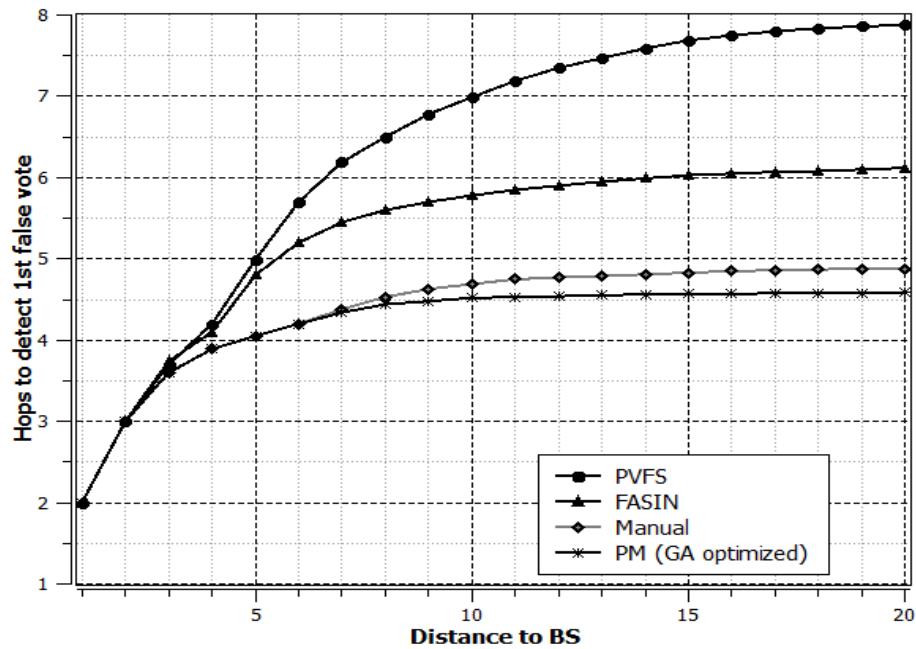


Fig. 8. Hops travelled by a report until first false MAC is detected.

5. Time Cost of Fuzzy Optimization

An exhaustive search method for the optimization of fuzzy membership functions may literally take several years usually in order of thousand years [17]. GA based membership-function optimization technique saves time cost and retrieves the optimal solution in much lesser time. The GAFPS finds an optimal solution for 9,000 trial sets (30 chromosomes \times 300 generations) in less than a day in the worst case i.e., if all the 9,000 trial sets are different from one another. Furthermore, during the evaluation of the generations, several identical chromosomes exist in a particular generation of chromosomes. Therefore further energy can be saved by evaluating only one of the identical chromosomes. Similarly, in the subsequent generations, several offspring chromosomes are identical to their parent chromosomes in the previous generations since the crossover and mutation probabilities are usually less than 1. The probability of having identical chromosomes in a single generation increases with an increasing number of identical chromosomes in a single population as the optimization process progresses. Therefore GAFPS finds an optimal solution within a feasible time for the network.

6. Conclusion

Genetic algorithms are an excellent tool for optimization of fuzzy membership function and require little human expertise. In this paper, we use GA based optimization technique to fine-tune the fuzzy membership functions in FAPS. GA based membership-function optimization technique saves time cost and retrieves the optimal solution in much lesser time. A chromosome represents the parameters that define the membership functions. Each chromosome in every generation is evaluated for its performance in a simulation unit. Genetic operators such as crossover, mutation are applied with their designed probabilities during the evaluation process and selection operator chooses the fittest chromosomes in the current population to be reproduced in the next population. Network energy, average hops traversed by

false reports are the two factors that are used in evaluating a chromosome for its fitness value. A significant amount of energy is saved by evaluating only one of the identical chromosomes. GA based optimization processes that make use of natural evolution methods, present a promising tool that supports the optimization of the parameters of fuzzy rule-based systems in WSNs.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. NRF-2015R1D1A1A01059484).

REFERENCES

- [1] L. Buttyán, L. Dóra, and I. Vajda, "Statistical wormhole detection in sensor networks," in *European Workshop on Security in Ad-Hoc and Sensor Networks*, pp. 128-141, 2005.
- [2] X. Han et al, "Fault-tolerant relay node placement in heterogeneous wireless sensor networks," *IEEE Transaction on Mobile Computing*, vol. 9, no. 5, pp. 643-656, 2010.
- [3] F. Ye et al, "Statistical en-route filtering of injected false data in sensor networks," *IEEE journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 839-850, 2005.
- [4] M. Akram and T. H. Cho, "Energy efficient fuzzy adaptive selection of verification nodes in wireless sensor networks," *Ad Hoc Networks*, vol. 47, pp. 16-25, 2016.
- [5] A. Wood, V. Srinivasan, and J. Stankovic, "Autonomous defenses for security attacks in pervasive CPS infrastructure," in *Proc. DHS: S&T Workshop on Future Directions in Cyber-Physical Systems Security*, 2009.
- [6] H. Y. Lee and T. H. Cho, "Fuzzy adaptive selection of filtering schemes for energy saving in sensor networks," *IEICE Trans. Commun.*, vol. 90, no. 12, pp. 3346-3353, 2007.
- [7] F. Li, A. Srinivasan, and J. Wu, "PVFS: a probabilistic voting-based filtering scheme in wireless sensor networks," *International Journal of Security and Networks*, vol. 3, no. 3, pp. 173-182, 2008.

- [8] Z. Yu and Y. Guan, "A dynamic en-route scheme for filtering false data injection in wireless sensor networks," in *SenSys*, 2005, pp. 294-295.
- [9] S. Zhu et al, "An interleaved hop-by-hop authentication scheme for filtering of injected false data in sensor networks," in *Proceedings of IEEE symposium on Security and Privacy*, 2004, pp. 259-271.
- [10] F. Ye et al, "A scalable solution to minimum cost forwarding in large sensor networks," in *Computer Communications and Networks, 2001. Proceedings. Tenth International Conference On*, 2001, pp. 304-309.
- [11] M. Akram and T. H. Cho, "Energy Efficient Fuzzy Adaptive Verification Node Selection-Based Path Determination in Wireless Sensor Networks," *Symmetry*, vol. 9, no. 10, pp. 220, 2017.
- [12] D. Ganesan et al, "Highly-resilient, energy-efficient multipath routing in wireless sensor networks," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 4, pp. 11-25, 2001.
- [13] J. Kim, Y. Moon, and B. P. Zeigler, "Designing fuzzy net controllers using GA optimization," in *Computer-Aided Control System Design, 1994. Proceedings., IEEE/IFAC Joint Symposium On*, 1994, pp. 83-88.
- [14] Z. Michalewicz, *Genetic Algorithms Data Structures= Evolution Programs*. Springer Science & Business Media, 2013.
- [15] K. H. Lee, *First Course on Fuzzy Theory and Applications*. Springer Science & Business Media, 200627.
- [16] S. Chi and T. Cho, "Fuzzy logic based propagation limiting method for message routing in wireless sensor networks," *Computational Science and its Applications-ICCSA 2006*, pp. 58-67, 2006.
- [17] H. Y. Lee and T. H. Cho, "Optimized fuzzy adaptive filtering for ubiquitous sensor networks," *IEICE Trans. Commun.*, vol. 94, no. 6, pp. 1648-1656, 2011.