LOAD BALANCING MANAGEMENT USING FUZZY LOGIC TO IMPROVE THE REPORT TRANSFER SUCCESS RATE

Sanghyeok Lim[1]and Taeho Cho[2]

[1]College of Information and Communication Engineering, Sungkyunkwan University, Republic of Korea
[2]College of Software, Sungkyunkwan University, Republic of Korea

ABSTRACT

A wireless sensor network (WSN) consists of multiple sensor nodes and base stations (BS) that collect information over widely deployed sensor nodes. Sensor nodes have limited energy source and low computing power. Due to those features, there is a disadvantage that user's individual node management is difficult and they are easily captured by attackers. Therefore, efficient energy allocation of nodes is important and network security protocol is needed. The Probabilistic Voting Filtering System (PVFS) is a system that prevents false vote injection attack and false report attack injected from attackers. The reason for the existence of this protocol is for energy management of nodes through defence against those attacks and in order to efficiently manage the network based on PVFS, load balancing of nodes should be performed. In the proposed scheme, fuzzy logic is applied to each cluster head node (CH) to perform load balancing by determine whether to perform a role as a verification node and an event forwarding node. The experiment shows that the event detection rate and the report delivery success rate are improved in proposed scheme compare with original PVFS.

KEYWORDS

Network Protocols, Wireless Sensor Network, Fuzzy logic, False Report Attack, False Vote Injection Attack

1. INTRODUCTION

WSNs are used for data collection and event detection in various fields such as home networks, military systems, and forest fire monitoring [1] and are composed of many sensor nodes and a base station (BS). When an event occurs, the sensor node detects the event and makes report of that event and send it over multiple hops of the sensor nodes to the BS[2]. However, sensor nodes are vulnerable to attack because of the disadvantages of limited computation, limited energy, random distribution in an open environment that operates independently, and difficulties in individual management [3], [4].Attackers exploit these vulnerabilities to attack WSNs by injecting reports containing false information or injecting false Message Authentication Codes (MACs). Figures 1 shows a schematic of these attacks. These attacks reduce the energy of the node, shorten the lifetime of the network, and prevent detection and reporting of normal events. Several protocols have been developed to prevent false report attacks[5],[6] andLi and Wu proposed a probabilistic voting-based filtering scheme (PVFS) [7]to prevent both false report and false vote injection attacks.

Natarajan Meghanathan et al. (Eds) : CCSEA, NCOM, AIFU, DKMP, EMSA, SIPRO, SEA - 2018
pp. 01– 13, 2018. © CS & IT-CSCP 2018 DOI : 10.5121/csit.2018.80301

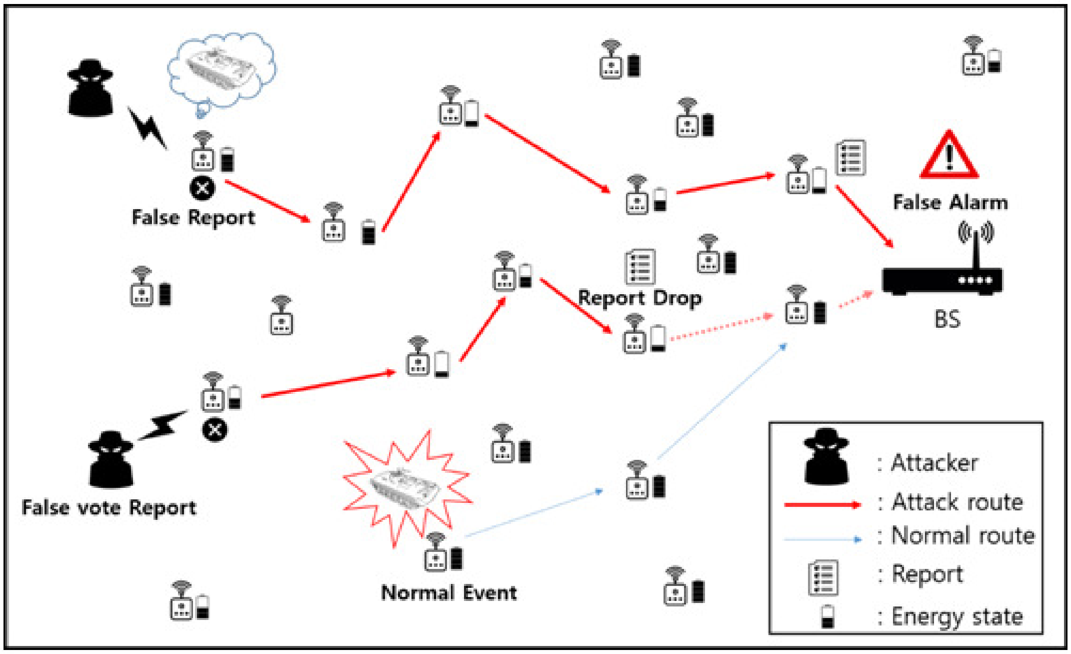

Figure 1. False report and false vote injection attacks

In PVFS, all nodes constitute a network that exploits cluster-based organization. When a cluster head (CH) recognizes an event, it generates a report of that event. It then sends this report to the member nodes. Next, the member nodes judge the authenticity of the report and generate their own MACs, alternatively referred to as votes in PVFS.

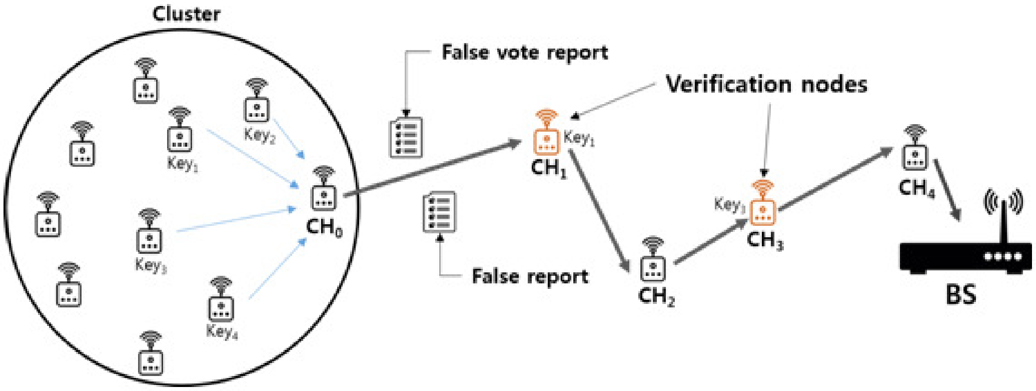

Figure 2. Report sending and verification process

Nodes that act as verification nodes in PVFS are more energy intensive than other member nodes, and if these nodes die, report generation and delivery of the region becomes impossible. In this paper, each CH's load balancing is performed by fuzzy logic based node role determination to increase event detection success rate and report transmission success rate with PVFS.Fuzzy logic is based on the fuzzy set thinking concept introduced by Professor L. A. Zadeh of the University of California at Berkeley in 1965 to quantitatively express the ambiguity of natural language and the like [8]. In the proposed method, fuzzy logic is used to help determine the role balancing of nodes by quantifying ambiguous properties such as how big the amount of nodes' energy is,how important nodes are.

2. RELATED WORK

2.1. PVFS

PVFS uses a true threshold value (Tt) and a false threshold value (Tf) to detect and filter false reports and false vote injected reports with validation nodes. PVFS has three phases; a key distribution phase, a report generation phase, and a verification phase.

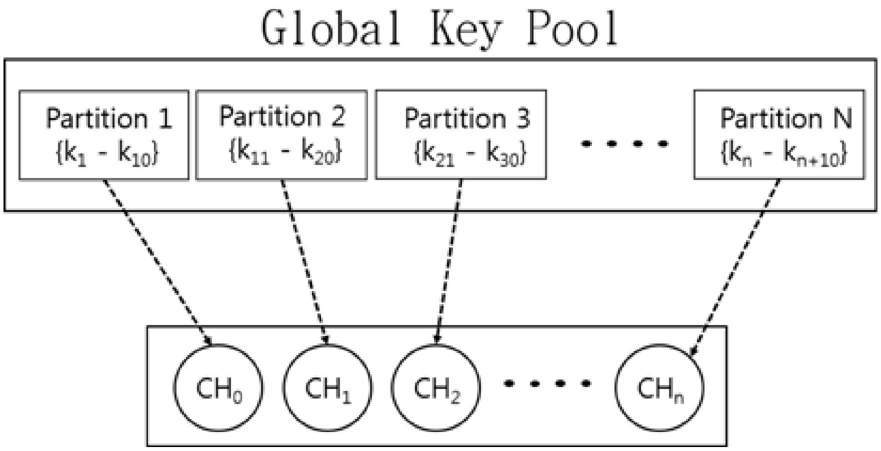
Global Key Pool
Partition 1 $\{k_1 - k_{10}\}$
Partition 2 $\{k_{11} - k_{20}\}$
Partition 3 $\{k_{21} - k_{30}\}$
Partition N $\{k_n - k_{n+10}\}$
CH_0
CH_1
CH_2
CH_n

Figure 3. Key distribution process

Figure 3 shows the key allocation step in which the BS divides the key pool into N partitions and delivers them to each CH. Each partition contains L keys. The CH uses one of the keys in the partition as its own key and distributes the remaining L-1 keys to it's member nodes. A key is allocated to the member nodes according to the partition of the key pool. With this process, every node gets a single key from a global key pool.

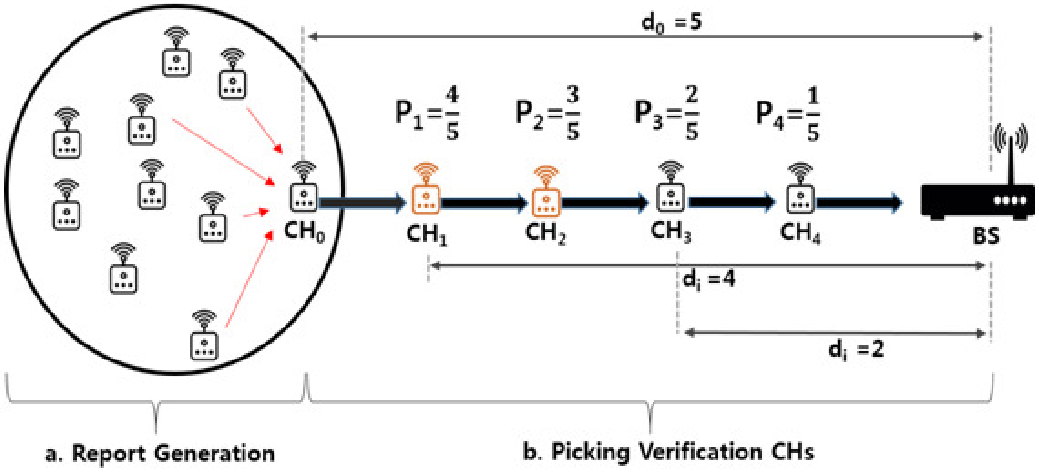

Figure 4. Report generation and verification node selection process

Figure 4 shows the report generation and validation node selection process. In the initial network configuration, nodes are divided into cluster units and the CHs responsible for report generation are selected ineach cluster. When an event occurs, the cluster closest to the event area wins the chance to generate a report of that event. CH generates a report and broadcasts it to member nodes of its cluster. The member nodes confirm this, and if the report is judged to be a normal report, the MAC created by its own key is transmitted to the CH. CH extracts a predetermined number of MACs received from the member nodes and adds them to the report. The node selected as the verification node stores the keys of the member nodes of the event occurrence cluster one by one. In the report verification process, the verifying nodes compare their own key indexes with that of in the report. If they have same key index, they extract the MAC through their key.. If the MAC generated by the same key is different, the vote is regarded as fake and Tf is increased. In the filtering process, if the false count reaches the threshold value, the report is judged to be false and is immediately dropped. If the true count value reaches the threshold value, the report is considered legitimate and is sent to the BS without further validation. The verification node selection process is shown in Figure 4-b. The verification nodes among the CHson the path to the BS are probabilistically selected to verify the report. The probability 'p' uses the distance d_0 which is the hopcount from the BS to the event cluster and the distance d_i between the BS and CH_i. The closer the CHs are to the event cluster, the higher the probability of verification.

2.2. Fuzzy logic system

Fuzzy logic is based on the fuzzy set thinking concept introduced by Professor L. A. Zadeh of the University of California at Berkeley in 1965 to quantitatively express the ambiguity of natural language and the like. [9], [10], [11]. The concept of a fuzzy set is a set out of binary logic that each object belongs to or belongs to a certain group and represents the degree to which each object belongs to the group as a membership function. Also, the fuzzy measure indicates the ambiguity of the fact that the ambiguous element a in the general set A belongs to the subset P of A, so that the relation between P and A is a mathematically continuous attribute. The configuration of the Fuzzy Logic Controller is shown in figure 5.

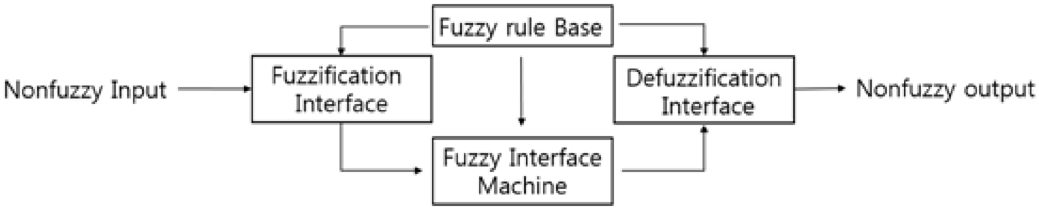

Figure 5. Fuzzy logic controller

The fuzzy controller consists of a fuzzy rule base, a fuzzy inference engine, and a defuzzifier. Fuzzification Interface is responsible for Scale Mapping task that changes the value range of the input variable to the universe of discourse. Fuzzy Rule Base is a set of linguistic control rules written in IF-THEN form. Fuzzy Inference Machine is decision-Making Logic that adopts rules from Fuzzy Input and Fuzzy Rule and in defuzzication Interface, scale mapping work is performed to convert into Universe of Discourse which is matched with the range of output variable value.

3. PROPOSED METHOD

3.1. Problem statement

In an environment where a WSN is installed, there are many areas where the user cannot distribute nodes directly and the nodes are scattered randomly. Therefore, there is a high possibility that a good routing path based on energy efficiency and characteristics of PVFS is not generated every time. One of the main purposes of the WSN is to gather information about events that occur in vast areas where it is difficult for a user to directly reach and respond appropriately. If the report of the detected event is not successfully transmitted to the BS, the main purpose of the WSN is to lose, or if the energy of the node is depleted so that event detection in the corresponding region is not possible. Also, if the residual energy of the sensor nodes is high but the events to be delivered are not properly transmitted, the sensor nodes distributed in those areas become useless. Therefore, it is important to improve the event detection rate and report delivery success rate even if the total energy consumption of the WSN increases by modifying the node to be available for those multiple purposes. Since PVFS shows probabilistic security, there is a tendency to show extremely high and low performance in random routing environments. Therefore, it is important to use nodes efficiently in the WSN by adjusting the role of nodes appropriately and raising the event detection rate. Figure 6 shows the part of the field where the WSN is configured. As shown in the figure, CHs that detect the event have limited communication distances, so they need to cooperate with other nodes to deliver the report to the BS through hop movement.

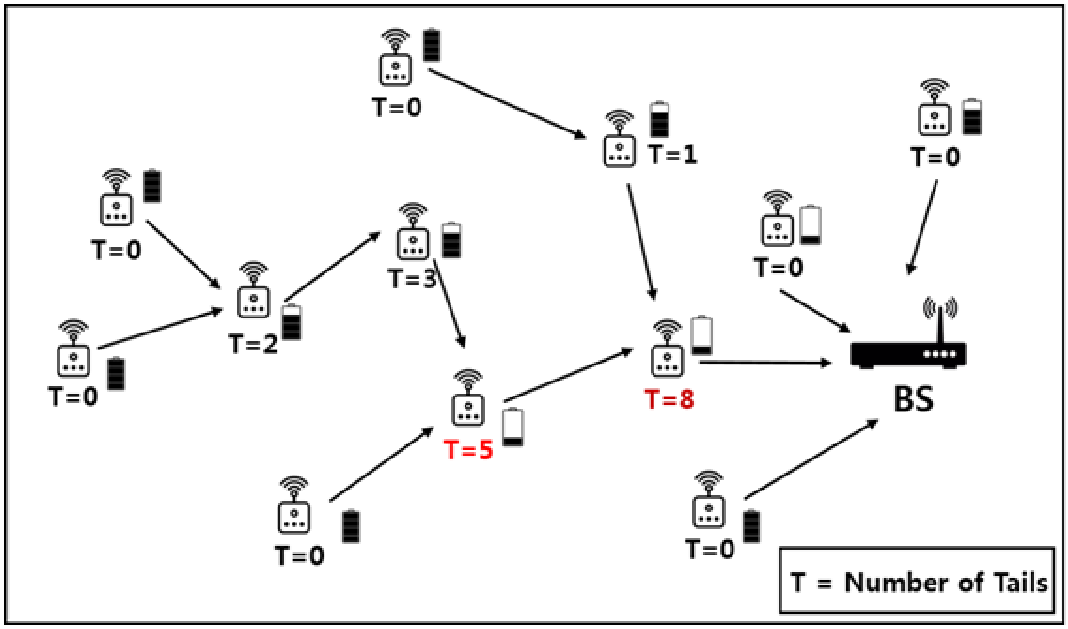

Figure 6. Calculating tail number

Therefore, each node has different number of downstream nodes depending on its location. Tail refers to the number of CH nodes that use the node to forward detected events. If the CH is positioned at the end of the field, the tail value of the corresponding node is zero. A node with many tails can consume a large amount of energy for sending, receiving, and verifying the report, which plays an important role in transferring report of the events occurring in other regions. If such node dies, the event detection node must forward the report through the new path except that node, and the event detected in the area where the dead node was receiving is no longer detectable because the CH is selected only once in the initial node dispatch phase.

3.2.System overview

Figure7 shows an example of the network on a field using the proposed method. Every CH has a list of the IDs of the nodes that are closer to the BS within its communication range. This list is used for the formation of new routing except the nodes selected by the tagging node among the list members. The sensor nodes have enough memory to store this brief information.

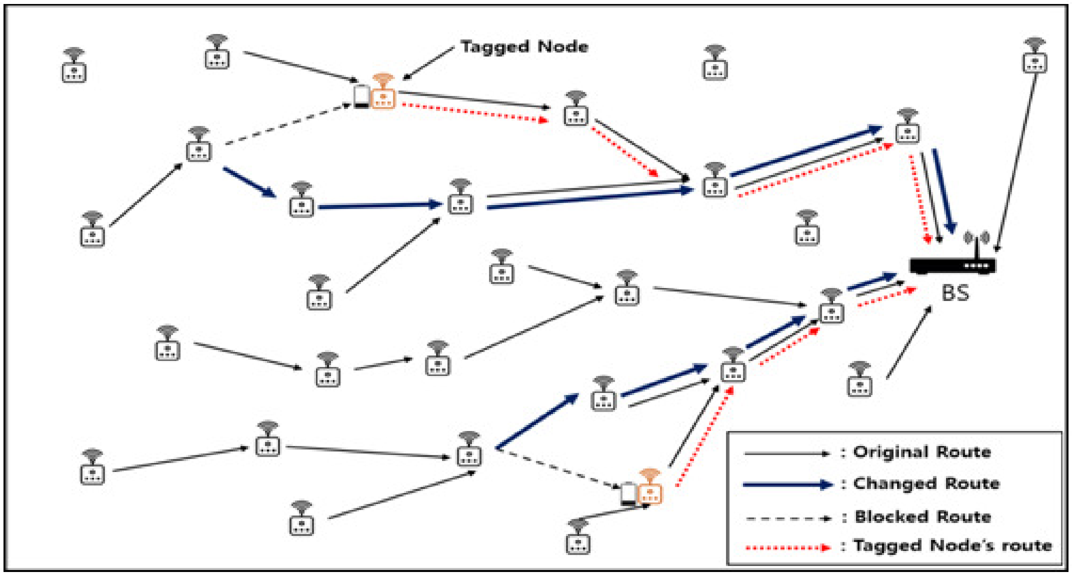

Figure 7. Dynamic routing in proposed scheme

Nodes with depleted energy are dropped from the list and the tagged nodes remain in the list but are excluded from the routing priority. If all the nodes in the list are tagged, the node closest to the BS is forcibly routed again. The node tagged by the fuzzy logic based decision system on the original route requests the previous route node to set up new routing, and the tagged node performs only event detection and report generation during the remaining period. The fuzzy function has three inputs: the energy of the node, the geographical importance of the node, and the frequency of occurrence of the event in the region. The calculation of the fuzzy function is performed with a constant period when it falls below a certain energy because frequent fuzzy calculations shorten the lifetime of the node.

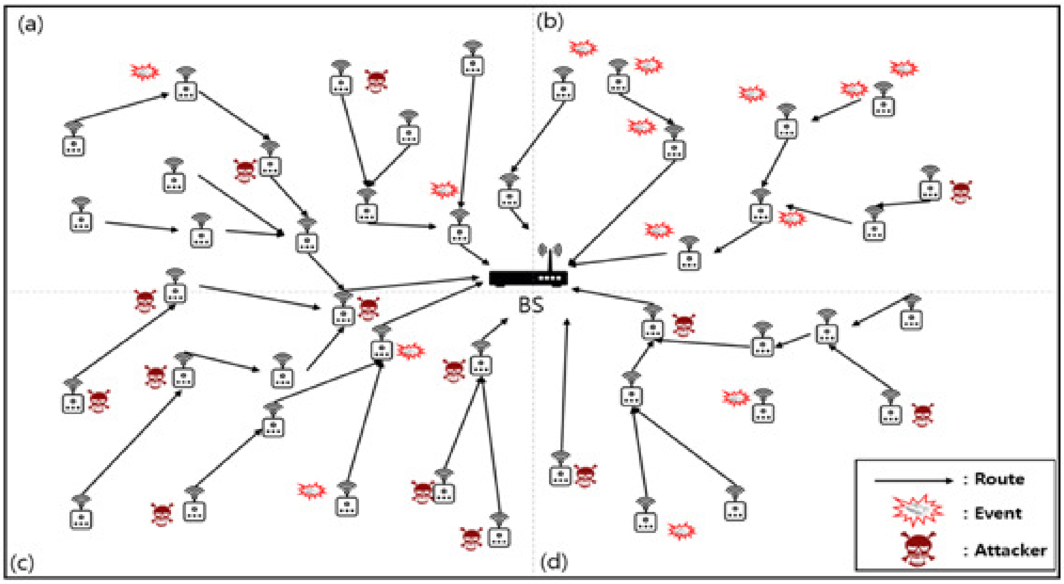

Figure 8. dynamic event and attacks in WSN

In the actual WSN field, there is very little chance that events and attacks will occur uniformly locally. In particular, in the case of an attack, it is difficult for the attacker to travel around the entire field, and attack occurs through a few nodes that have been hijacked. Therefore, load balancing considering the local attack rate is necessary. This is also why the frequency of occurrences of fuzzy input values in the proposed method is getting into.Figure 8 shows how these attacks and normal events can occur in an actual field.The proposed scheme extend the lifetime of the network through load balancing of nodes considering these characteristics, and improve report transmission success rate and event detection rate.

3.3. Fuzzy based fitness evaluation

The proposed method exploits a fuzzy rule based system for load balancing. Part of the appeal of fuzzy rule based systems is that they can be used for approximate reasoning. Which is particularly important when there is uncertainty in reasoning processes in addition to imprecision in data. The fuzzy input includes the amount of residual energy of the node, the importance of the position of the node, and finally the frequency of local event occurrence. The fuzzy output for these three inputs helps determine whether to use the CH as an event delivery and verification node.

3.4. Fuzzy membership function

The functions for the three inputs of the fuzzy logic are shown in the figure 9.The following is the reason for selecting the values used in the fuzzy input.The fuzzy functions contain the importance, energy amount, and frequency of occurrence of event and attacks respectively. Optimization of functions can be performed using genetic algorithms[12],[13],[14].In the proposed method, several functions are randomly simulated and a function that most satisfying the if-then rule is used.

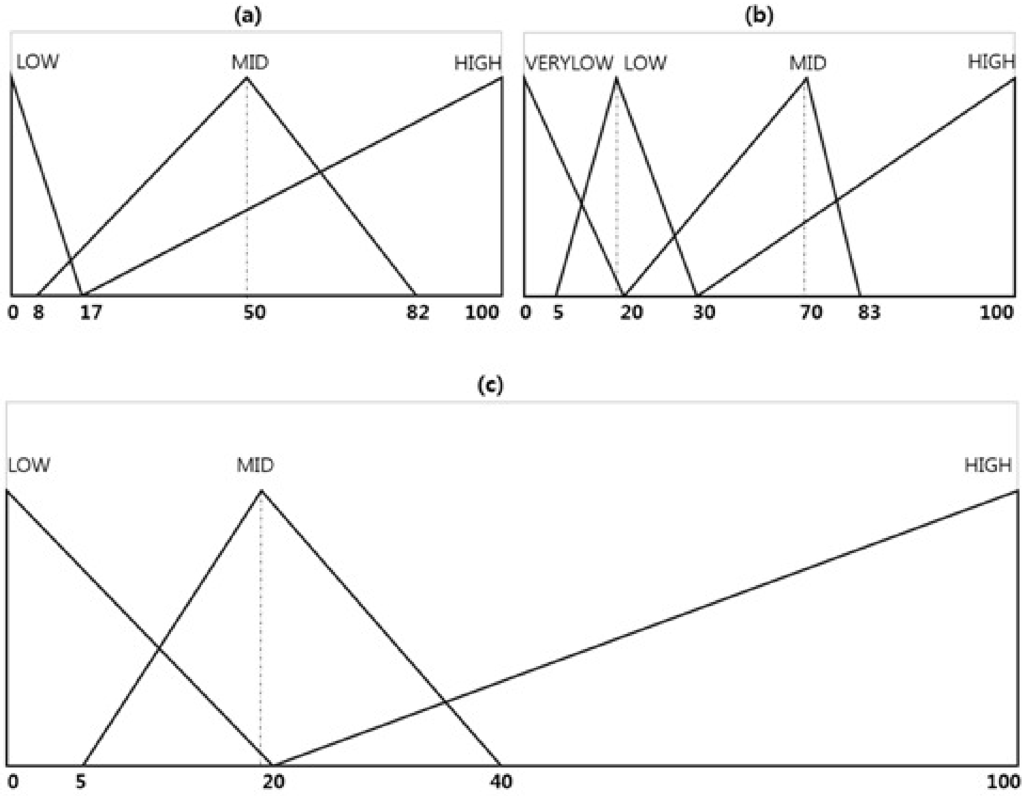

Figure 9. Fuzzy membership functions

1. Importance (a): Importance is a measure of whether a given CH plays an important role in the network. It is a value considering the probability that the CH becomes the verification node of the downstream nodes and the tail number of the corresponding CH.A node with a high probability of being a verification node is expected to have a high energy consumption and its importance is increased. In general, in PVFS, the closer to BS, the more tail increases and the probability of becoming a verification node becomes lower. Therefore, the distance with BS does not lower the importance of corresponding nodes every time and it's hard to measure it just considering distance. Therefore, the proposed method uses the measurement function to measure the importance. The evaluation function of i-th node E_{tag}^{i} is expressed as the sum of α, β and γ, which are the amount of energy used during the role of the forwarding node, the amount of energy consumed during the role of the event detection node and the energy used by the verification node to report transmission after event detection, respectively. α is given by $\alpha = e_t * \frac{N_{evt}}{N_n}$, where e_t is the transmitting cost, N_n is the number of nodes in the field and N_{evt} is the number of events. β is given by $\beta = T * (e_t + e_r) * \left(\frac{N_{evt}}{N_n}\right)$, where e_r denotes the receiving cost and *T* is the *i*-th CH's tail node. γ is more complicated, and can be expressed as the sum of probabilities that the i-th node is a verification node of each tail node, all times the probability of having the same key when it is a verification node. The sum of probabilities that a node is a verification node of the tail nodes is given by $e_{cal} * \sum_{j=1}^{n} P_i^j$, where $P_i^j = \frac{HopCount_{N_j}}{HopCount_{N_i}}$ and e_{cal} denotes calculating cost.Then, the sum of probabilities of each node is $P_i^1 + p_i^2 + \cdots + P_i^T = \frac{1}{HopCount_{N_i}} * \sum_{j=1}^{T} HopCount_{N_j}$. When the node performs verifying, the probability that the key of the corresponding node overlaps with the key of the report is$\frac{S}{L}$, where s is the number of votes in the report and L denotes the number of member nodes from each tail nodes. Therefore, the evaluation function is as follows:

$$E_{tag}^{i} = e_t * \frac{N_{evt}}{N_n} + T * (e_t + e_r) * \left(\frac{N_{evt}}{N_n}\right) + \frac{S}{L} e_{cal} * \frac{1}{HopCount_{N_i}} * \sum_{j=1}^{T} HopCount_{N_j}.$$

The output of the function is integerized into the fuzzy input.

2. Energy (b): The energy of the node is the most important element of the fuzzy inputs. When nodes with sufficient energy are tagged, the routing path is unnecessarily increased, which has the adverse effect of increasing transmission and reception costs. Also, the probability of encountering the hijacked node during the hop movement increases. On the contrary, if the CH is tagged in a situation where the amount of energy is too small, there is not enough residual energy to perform an event detection function, so that a node is killed during the generation of a report or waiting for a report, Report passing rates are also lower.

3. Event occurrence frequency (c): The occurrence frequency of events indicates the ratio of events occurring in the area where the corresponding node is located among all the fields. If the event frequency is high, the probability of event detection and report generation increases, and the probability of being tagged increases accordingly. On the other hand, if the frequency of occurrence of events in the area is low, it is desirable from the viewpoint of the whole network to delay the tagging timing and induce the node to perform the role of the verification and transmission node for a longer time. The frequency of occurrence is difficult to measure in real time, and the amount of energy consumed in real-time measurement increases, which is rather counterproductive.

The following is if-then rule of fuzzy system. It contains those 3 inputs that previously explained. Each input values have 3, 4 and 3 levels, respectively and it's the output of the rule is determined by the expert.

Table 1. Fuzzy If-Then rules

No.	INPUT			OUTPUT
	Importance	Energy	Event occurrence	ON/OFF
1	LOW	VERYLOW	LOW	ON
13	MID	LOW	LOW	OFF
15	MID	LOW	HIGH	ON
18	HIGH	LOW	HIGH	ON
31	MID	HIGH	LOW	OFF
36	HIGH	HIGH	HIGH	OFF

Table 1 shows some of the 36 if-then rules used in the proposed fuzzy system. The 3 input values are Importance, Energy, and event occurrence, and the output contains whether or not to tag the node as a report-generating-only node. For example, if the importance is LOW, the energy is VERYLOW, and the event occurrence is LOW, the node is selected as a node for event generation only. If the importance is MID, energy is LOW and event occurrence is LOW, It is decided not to tagged yet. The input can't be arbitrarily adjusted, and it must have a total of 36 values. The output of the rule reflects the case with the best results through experiments with several combinations. Rule with inconsistent output results in system performance degradation, so the exorbitant output was excluded from the experiment.

4. EXPERIMENTAL RESULT

4.1. Experimental environment

Table 2. Experiment parameters

Item	Value
Sensor field size(m)	1000× 1000
Number of sensor nodes	2000
Number of cluster head nodes	200
S	5
L	10
Packet size(byte)	24
Transmission range(m)	150

Experiments were conducted assuming that the attack rates were 0 to 90% The reason a 100% attack rate is excluded from experiments is that this experiment is an experiment that evaluates the failure rate of a normal report. The transmission and reception costs are set to 16.25μj and 12.5μj, respectively, and the calculated cost of voting was set to 15μj [15].The threshold values Tt and Tf used in the experiments were in accordance with the experimental environment of the original PVFS[7].The reason for this is to increase the reliability of the proposed method by making the environment of original version of PVFS equal to the proposed method.

4.2. Assumptions

The CH's location and key distribution phase does not change during the event. If the initial energy of the node is too high, it will be difficult to calculate the event detection failure and report transmission failure accurately because the node's energy depletion does not occur during the experiment. Therefore, the initial energy of the node is set to an appropriate amount. The BS has all the keys distributed to the node and has the computing power to verify all false reports, false votes.

4.3. Experimental result

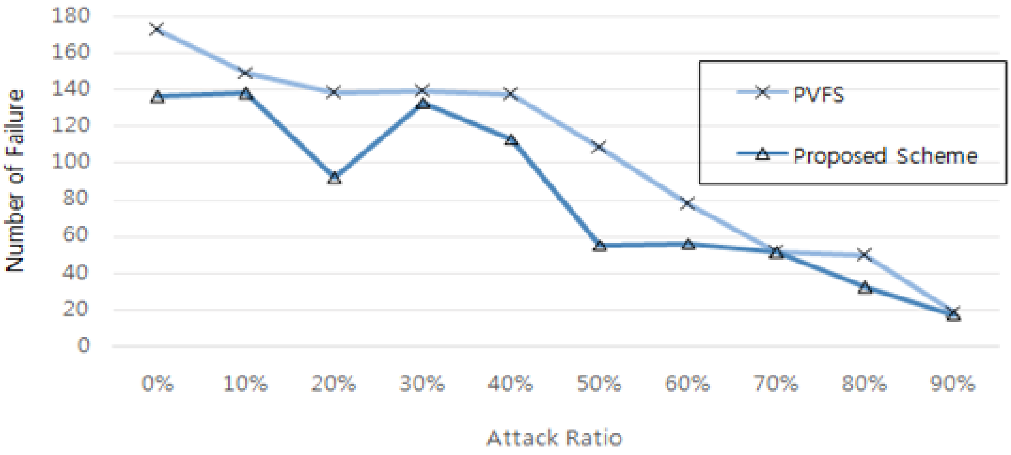

Figure 10. Number of transfer Failure

Figure 10 shows the number of report failures according to the attack rate at 600 events when the local occurrence frequency is 20, 40, 30, and 10%.The experiment was conducted with the exception of the transmission success rate for false report attacks because the network users are not mean to receive false reports derived from the attacker. It can be seen that the number of report transmission failures of the existing PVFS and the proposed technique decreases as the attack rate increases because when an attack ratio increases, the number of normal event's report sending itself is decreases. And as the attack rate increases, it can be seen that the difference between the number of report transmission failures of the proposed technique and PVFS is also reduced.

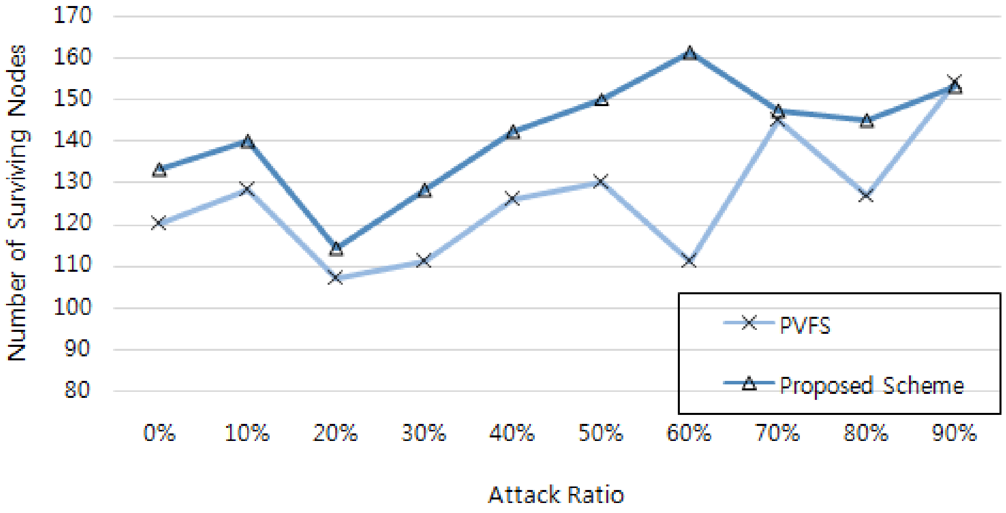

Figure 11. Number of surviving Nodes

Figure 11 shows the number of surviving nodes according to the attack rate of the existing PVFS and the proposed scheme. The number of surviving nodes was measured at every cycle after the experiment, and there is some part with little difference from original version of PVFS (70%, 90%).The reason for this result is that the proposed scheme has more number of surviving nodes than the existing PVFS during the experiment, resulting in additional energy consumption due to event reception, report generation, and report transmission. These results show that the number of failures of the proposed method and the report delivery failure of the existing PVFS are significantly different from those of the Figure 10, but the number of surviving nodes is not so different.

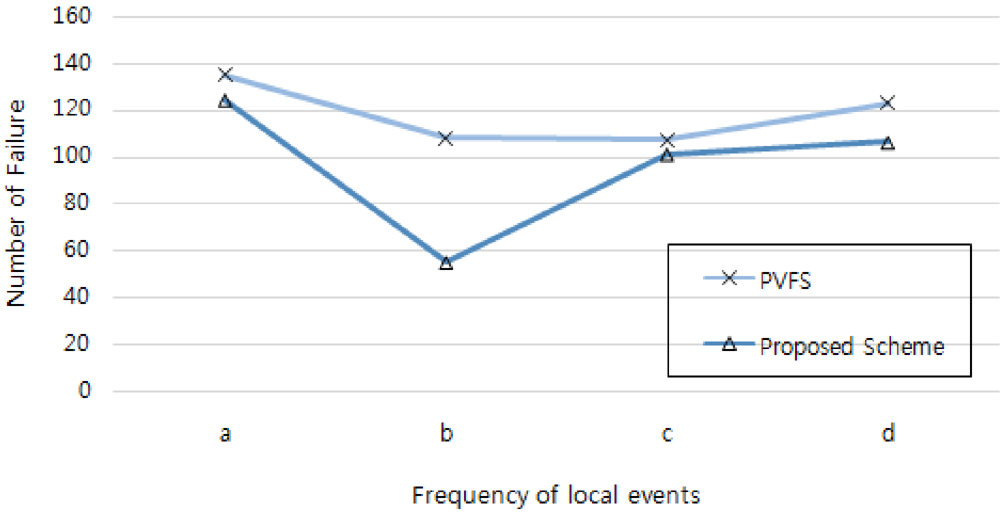

Figure 12. Number of Transfer Failure in Dynamic WSN Field Environment

Figure 12 shows the performance difference between the proposed method and the existing PVFS according to the change of local incidence a, b, c, and d of Figure 12 represent the frequency of events occurring in the regions of Figures 8-a, 8-b, 8-c, and 8-d, respectively and each variable means a = 25,25,25,25(%), b = 30,40,20,10(%), c = 20,20,50,10(%), d = 0,50,0, 50(%).As can be seen from the graph, load balancing with fuzzy logic increases the lifetime of the entire network and increases the success rate of the normal report on the event. A large number of surviving nodes means that event detection for the region can be longer and more. The residual energy of all the nodes can be measured somewhat lower in the proposed method because the report about the event that occurred in the dead node is generated and transmitted, the event detection is not performed, and report transmission more.

3. Conclusions

In the proposed method, we experimentally confirmed that a WSN's event detection success rate, report transmission success rate, and overall network lifetime are increased by organically changing the roles of the nodes through fuzzy logic based decision making system that evaluate the node status. What the user wants through the WSN is successful reporting of normal events. Therefore, the experiment of the proposed technique excluded how successful the transmission of false reports was. The WSN user periodically replaces the node battery placed in the entire field or picks up the nodes placed in the field when the local network is not needed. Therefore, the total sum of the residual energy amount of the entire nodes placed in the field does not influence the performance of the network. Although it is relatively low energy, it should be the criterion for determining the performance of the network whether or not the node can generate a report on the live event and send it to BS. The proposed method is a method of raising the report success rate and the number of surviving nodes simultaneously. For further study, the load balancing using the membership function of the fuzzy logic which is optimized through the genetic algorithm will be implemented. The routing environment changes every moment and it is impossible for humans to optimize the membership function at every moment, so the genetic algorithm should be applied to automatically adapt the changing routing environment to obtain the optimal output.

ACKNOWLEDGEMENTS

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] Zhang, Wensheng, and Guohong Cao. "Group rekeying for filtering false data in sensor networks: A predistribution and local collaboration-based approach." INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE. Vol. 1. IEEE, 2005.

[2] Perrig, Adrian, John Stankovic, and David Wagner. "Security in wireless sensor networks." Communications of the ACM 47.6 (2004): 53-57.

[3] Al-Karaki, Jamal N., and Ahmed E. Kamal. "Routing techniques in wireless sensor networks: a survey." IEEE wireless communications 11.6 (2004): 6-28.

[4] Perrig, Adrian, John Stankovic, and David Wagner. "Security in wireless sensor networks." Communications of the ACM 47.6.

[5] Zhu, Sencun, et al. "An interleaved hop-by-hop authentication scheme for filtering of injected false data in sensor networks." Security and privacy, 2004. Proceedings. 2004 IEEE symposium on. IEEE, 2004.

[6] Yang, Hao, and Songwu Lu. "Commutative cipher based en-route filtering in wireless sensor networks." Vehicular Technology Conference, 2004. VTC2004-Fall. 2004 IEEE 60th. Vol. 2. IEEE, 2004.

[7] Li, Feng, and Jie Wu. "A probabilistic voting-based filtering scheme in wireless sensor networks." Proceedings of the 2006 international conference on Wireless communications and mobile computing. ACM, 2006.

[8] Zadeh, Lotfi A. "Fuzzy sets." Information and control 8.3 (1965): 338-353.

[9] J. Yen and R. Langari, Fuzzy Logic: Intelligence, Control, and Information. Prentice-Hall, Inc., 1998.

[10] G. Klir and B. Yuan, Fuzzy Sets and Fuzzy Logic. Prentice hall New Jersey, 1995.

[11] R. Babuška, "Fuzzy Systems, Modeling and Identification," Delft University of Technology, Department of Electrical Engineering Control Laboratory, Mekelweg, vol. 4, 1996.

[12] D. E. Golberg, "Genetic algorithms in search, optimization, and machine learning," Addion Wesley, vol. 1989, 1989.

[13] C. L. Karr, "Design of an adaptive fuzzy logic controller using a genetic algorithm." in Icga, 1991, pp. 450-457.

[14] A. Geyer-Schulz, Fuzzy Rule-Based Expert Systems and Genetic Machine Learning. Physica Verlag, 1997.

[15] Ye, Fan, et al. "Statistical en-route filtering of injected false data in sensor networks." IEEE Journal on Selected Areas in Communications 23.4 (2005): 839-850.

AUTHORS

Sanghyeok Lim

received a B.S. degree in Digital Information Engineering from Hanguk University of Foreign Studies in 2017, and is now working toward an M.S. degree in the Department of Electrical and Computer Engineering at Sungkyunkwan University.

Taeho Cho

received a Ph.D. degree in Electrical and Computer Engineering from the University of Arizona, USA, in 1993, and B.S. and M.S.degrees in Electrical and Computer Engineering from Sungkyunkwan University, Republic of Korea, and the University of Alabama, USA, respectively. He is currently a Professor in the College of Information and Communication Engineering, Sungkyunkwan University, Korea.