

GEOGRAPHIC INFORMATION-BASED ROUTING OPTIMIZATION USING GA FOR CLUSTER-BASED WSNS

Sanghyeok Lim¹ and Taeho Cho²

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

²College of Software, Sungkyunkwan University, Republic of Korea

ABSTRACT

Wireless sensor networks are used for data collection and event detection in various fields such as home networks, military systems, and forest fire monitoring, and are composed of many sensor nodes and a base station. Sensor nodes have limited computing power, limited energy, are randomly distributed in an open environment that operates independently, and have difficulties in individual management. Taking advantage of those weaknesses, attackers can compromise sensor nodes for various kinds of network attacks. Several security protocols have been proposed to prevent these attacks. Most of the security protocols form routings with cluster head nodes. In the case of routing using only cluster head nodes, it is difficult to re-route when the size of the cluster is increased or the number of the surviving nodes is reduced. To prevent these attacks, the proposed scheme maintains security in a cluster-based security protocol and shows energy efficient routing using genetic algorithm by selecting the appropriate cluster head nodes and utilizing the characteristics of the sensor node with different transmission outputs based on the distance between each node. In this paper, we use a probabilistic voting-based filtering scheme, one of the cluster-based security protocols, and the shortest path, which is a hierarchical routing protocol that the original probabilistic voting-based filtering scheme is using, to test the proposed scheme. This experiment shows the performance comparison of the routing success rate and routing cost according to the number of nodes on the field, as well as the performance comparison according to the cluster size per number of nodes.

KEYWORDS

Network Protocols, Wireless Sensor Network, Network attack, Genetic algorithm, Routing

1. INTRODUCTION

Wireless sensor networks (WSNs) are used for data collection and event detection in various fields, such as home networks, military systems, and forest fire monitoring, and are composed of many sensor nodes and a base station (BS) [1], [2]. When an event occurs, the sensor node detects the event, makes a report of that event and sends it over multiple hops of the sensor nodes to the BS. 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. An attacker exploits these weaknesses to compromise nodes and perform various kinds of attacks. These attacks reduce the energy of the node, shorten the lifetime of the network, and prevent the detection and reporting of normal events. Several protocols have been proposed to increase the security and lifetime of the network by defending against such attacks [3], [4], [5], [6], [21]. These protocols form cluster-based routing, which benefits from local management of nodes and it makes systemic report generation possible. However, such cluster-based routing, which relies only on CH nodes for report transmission and verification, has a significant negative impact on security considering the entire

network because of various attacks, such as sniffing or dropping reports transmitted from the downstream event detection cluster. Additionally, considering the shortest path that is applied to most WSN fields, the node with the shortest distance to the BS and within the transmission range based on the report delivery node is selected as the next transfer node. When a sufficient number of nodes is appropriately placed in the field, this ensures stable routing formation and a reasonable cost.

However, if the number of nodes placed in the field is insufficient or the size of the cluster is inevitably increased due to a change in the field environment, the forwarding node candidate list is likely to be reduced, and it becomes difficult to guarantee energy efficient routing. This routing increases the total number of routing hops, thereby reducing the lifetime of the network. In order to prevent such problems, the CHs of each event detection cluster perform only event detection, report generation, and report transmission. The verification node is considered and selected among all the nodes with an appropriate geographical advantage considering the distance to the BS and transmission power. The existing layer-based routing schemes applied to most of the security protocols implement routing sets maximum transmission output without regard to the distance between nodes [7], [8]. Micaz mote used in most of the security protocols can set different transmission output levels according to each node's distance [9]. In the proposed scheme, the routing is constructed using genetic algorithm (GA) to minimize the cost of the routing by setting the appropriate output level for all nodes forming the routing. In addition, to minimize the distance of the receiving node from the initial event detection node, the CH node was selected as the node with the smallest distance from the BS, thereby saving the initial routing cost while existing protocol randomly select the CH or select the node with the smallest ID value. Through experiments, we confirmed that the performance difference between the proposed scheme and the existing scheme increases as the number of nodes decreases. We experimented our propose scheme based on PVFS which is one of a cluster-based security protocol using the shortest path routing.

2. RELATED WORK

2.1. PVFS

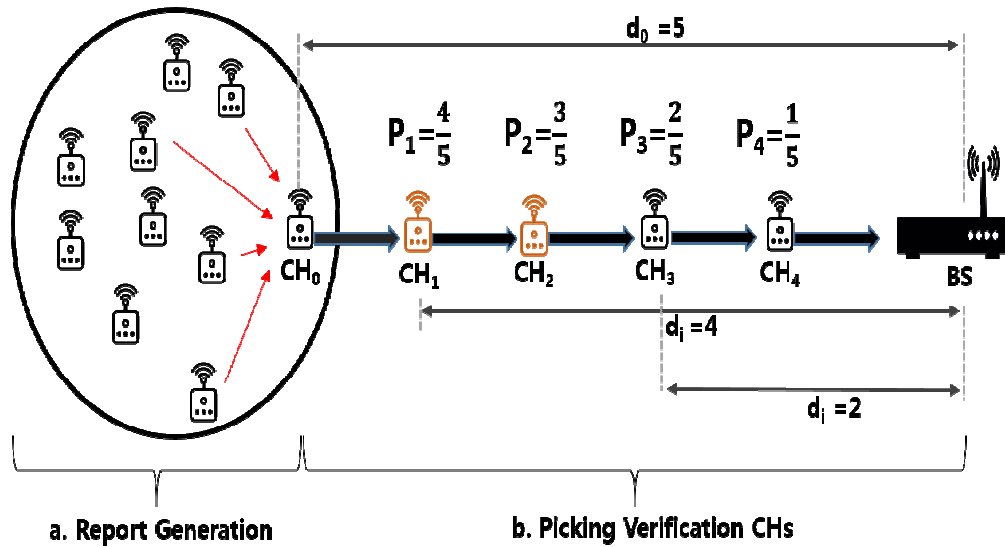


Figure 1. PVFS's en-route filtering.

PVFS is a typical cluster-based security protocol. Nodes randomly placed on the field are assigned to the same cluster within an appropriate range to detect events within the same cluster area. At the end of the node deployment, PVFS performs network security and report transmission in four steps.

- Key distribution phase

The BS transmits to each CH a set of keys separated by a cluster size. The CH that receives the key set distributes one key to each member node.

- Selection step of the verification node

The verification node is selected by the probability based on the number of hops between the event detection cluster and the BS, and the number of hops between the node and BS among the CHs on the path. The CH selected as the verification node receives one of the keys of the member nodes of the event detection cluster at random. The CH of the selected cluster in the verification node clears all keys except its own in its own key set.

- Report delivery phase

After detecting the event, the CH generates a report on the event, receives the MAC created by the member nodes of each node, and attaches it to the report to generate the final report.

- Verification step

When the verification node receives the report, it first compares its own key index with the key index of the MAC attached to the report. If the key index overlaps, the MAC check is performed. If the index does not overlap, the report is not verified and sent to the node on the next path. If a normal MAC is detected as a result of the check, a normal vote is cast. If a false MAC is detected, a false vote is cast. If the number of votes reaches a preset threshold value, the report is discriminated as normal or false. A report that is deemed to be a normal report is then immediately directed to the BS without any further verification steps.

2.2. Genetic Algorithm

The genetic algorithm (GA) was proposed by Holland John and takes ideas from evolutionary processes occurring in the natural world. This kind of algorithm represents solutions to a problem as bit strings and finds an optimal solution through evolution [10], [22], [23]. The GA unit probabilistically selects two individuals constituting the current generation. The selection probability for each individual is proportional to the fitness of the individual, and the child generations consist of individuals with greater fitness than the members of the parent generation. Selected individuals generate a new generation through child generating, mutating, and sometimes elite choosing [11], [12]. To solve the problem of local fixation in evolution the proposed GA uses mutation technic.

3. PROPOSED METHOD

3.1. Problem statement

3.1.1 Routing problems

A WSN with a vast land area has few obstacles between deployed nodes, and cluster-based security protocols mainly use the shortest path based routing. In a highly selective environment,

this is efficient for the forwarding node in the routing formation. However, in many security protocols, including PVFS, the selection range of the forwarding node is limited to the CH among all nodes. Moreover, once the CH is selected, it will not be replaced by any other member node. In the situation where the CH is being used for extortion or because of battery discharge, the corresponding routing scheme shows a significant weakness in re-routing. In some cases, when the next forwarding node is set based only on the distance to the BS, the selection of the geographically optimized node cannot be performed, which results in the adverse effect of the increased number of routing hops.

3.1.2 Delivery node selection problem within cluster-based security protocols

Most of the cluster-based protocols select only those nodes that are responsible for report delivery among the CH nodes. There are two bottlenecks in the routing path: key storage and energy consumption. The key storage bottleneck is a phenomenon in which the number of keys owned by the CH for the verification of downstream nodes increases. When these CHs are compromised, the attacker will have multiple keys and can cause a significant security problem. The bottleneck of energy consumption is the phenomenon that the routing of the downstream nodes to the BS proceeds through the corresponding CH. As the number of reports to be transferred to the CH increases exponentially, the expectation of energy consumption can be greater than other nodes and the battery will be discharged soon. The network will then have to go through the routing again and additional energy consumption is expected.

3.1.3 Uniform output power of sensor nodes

In the existing security protocols, all nodes transmit the report to its upstream nodes with a maximum transmission range. In actuality, it is unnecessary to communicate only with the maximum transmission distance by disregarding the interval between each sensor node. It is therefore useful to measure the distance between nodes and transmit them at the appropriate output level. The distance measurement between nodes can be implemented using the RSSI value [13], or the node attached GPS module [14]. The Micaz node, which is often used in WSNs, is a node that can control the output of a total of 32 levels, allowing the user to set the appropriate output value [15], [16], [17]. There are many experiments on the transmission distance according to the output in actual situations [18], [19], [20].

3.2. System overview

The experiment of the proposed method constitutes the system based on PVFS. All key distribution and en-route filtering procedures follow existing security protocols. After the node is deployed, each node sends its location information to the BS. Based on this information, the BS selects the nodes with the shortest distances from the BS itself in each cluster as CH. The difference in geographical location in the selection of CH has no effect on the security of the protocol. This is because in the existing security protocol, CH selection was made without taking any other factors into account. The key distribution and verification node selection phase follows the original protocol's phase. Then, the intermediate delivery node is randomly selected among all the nodes, including the cluster head node. The selection of the next forwarding node is based on the nodes within the maximum transmission range of itself and re-divides the range based on the distance to the BS among the nodes within its transmission range.

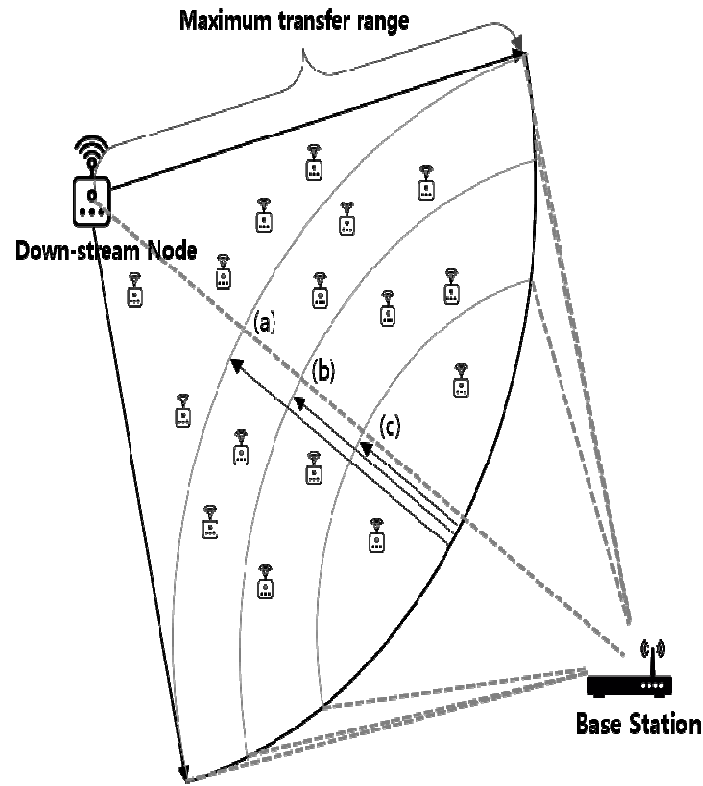


Figure 2. Range of the candidate list of the forwarding node.

In Figure 2, (a), (b), and (c) are the ranges of the candidate list of the forwarding node. More nodes can be selected by the downstream node with a larger range, resulting in an increase in the diversity of the nodes forming the routing. However, this means that there is less genetic information overlapping between parent chromosomes in the GA, and it is also difficult to create a next generation node having a cross-point between parents. The contents of the candidate list range of the forwarding node will be explained after dealing with the routing through the GA.

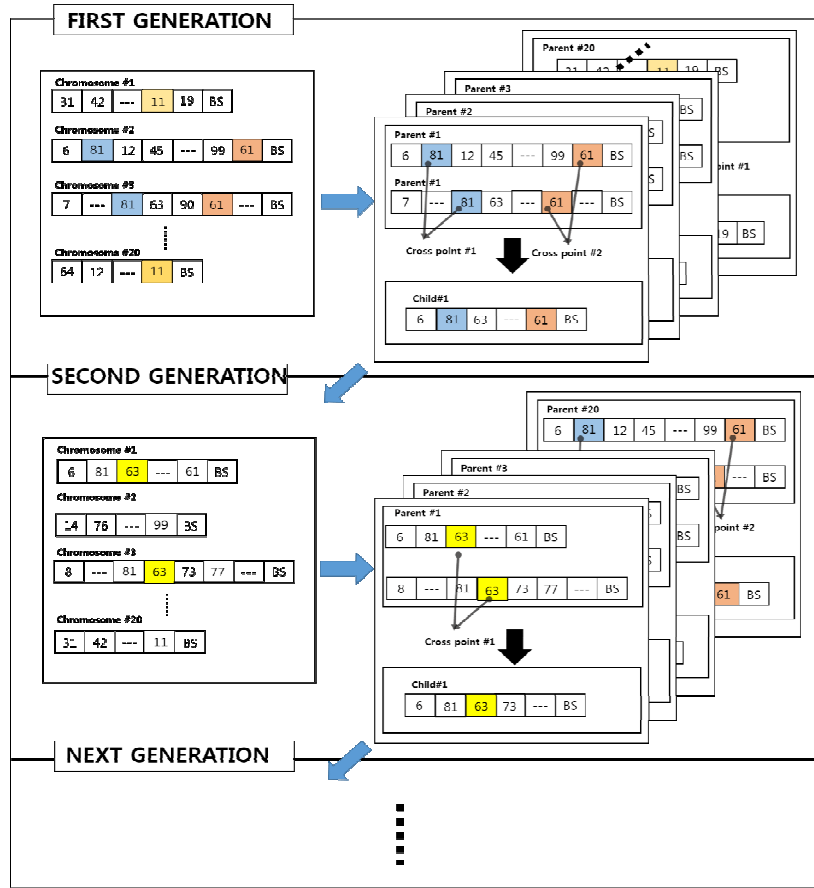


Figure 3. Routing optimization using the GA.

The GA chromosome contains the IDs of the forwarding nodes, and the length of each chromosome is equal to the number of hops in each routing path. The fitness of the chromosomes is inversely proportional to the final routing cost of the chromosomes. If the genetic information of the two parental chromosomes selected do not overlap, they will not have a cross-point. In this case, the parent chromosome is re-selected through the same method. If the candidate to be the parent chromosome is large, the probability that the genetic information of each other does not overlap increases. The proposed scheme limits the re - election of parental chromosomes until genetic information overlaps. When the number of re-elections reaches a predetermined threshold, chromosomes with the highest fitness are re-elected and transferred to the next generation. This is a modified version of the elitism technique used in genetic algorithms. Instead of mutation, we changed the size of the node list to increase the diversity of chromosomes.

Additionally, we did not use the elitism technique that causes the fixation phenomenon. The size of the candidate list of the transfer node of Figure 2 has a trade-off between the child chromosome generating ability of the GA and its routing cost. Therefore, if the network user has to re-select the routing within a short time, the scope of the candidate list can be narrowed and the number of GA routines can be reduced to form an immediate routing, even if the cost is slightly greater. Conversely, if the user wants to optimize the routing cost, the user can increase the range of the list and increase the number of repetitions of the GA.

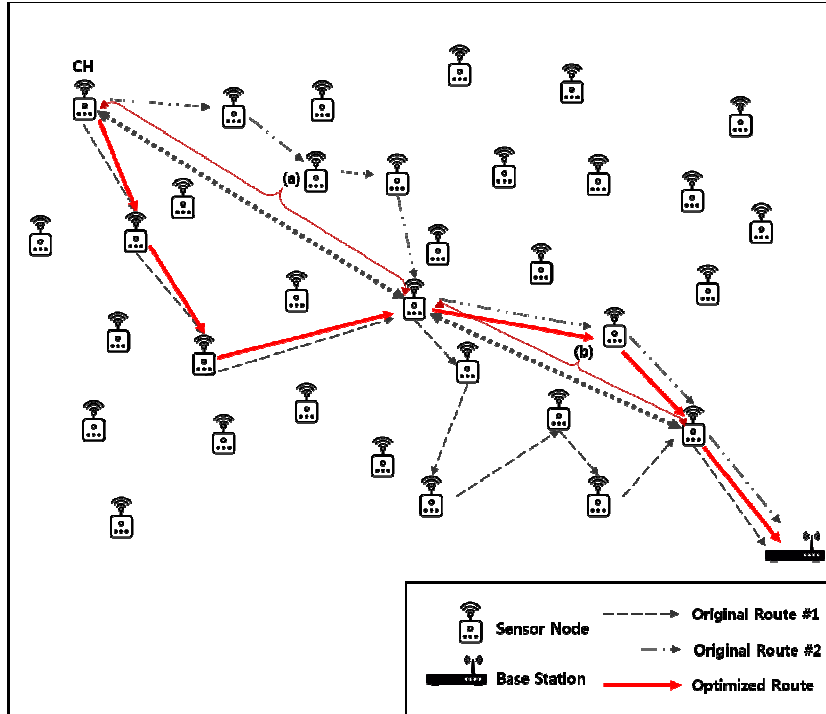


Figure 4. Route comparison.

Figure 4 shows the original routing and modified routing using proposed scheme. Route # 1 and Route # 2 have two cross-points. These are routing information for Parent # 1 and Parent # 2, respectively. Thus, the child chromosome starting from Route # 1 changes the path of the parent chromosome at the cross-point and forms the routing. In Section A, Parent #1 routing has fewer hops and less energy usage, and in Section B, Parent # 2 routing is more efficient. As the generation is repeated, the child route picks the optimal interval of the parent appropriately and eventually has less routing costs and higher expectations than the parent generation in this way and routing is then completed.

4. EXPERIMENTAL RESULT

4.1. Experimental environment

Table1. Experiment parameters

Item	Value
Sensor field size (m*m)	1000 × 1000
Number of sensor nodes	1000 - 2000
Cluster size	40 - 60
Transmission range (m)	80

4.2. Assumptions

The transmission and reception costs are set to 16.25µj and 12.5µj, respectively, and the calculated cost of voting was set to 15µj. In our experiment, Micaz motes were used as sensor nodes, which can select 32 distinct output levels [9]. Experimental results show that the proposed

method does not adjust the maximum output level as the conventional method does, but instead the proposed method adjusts the maximum output level. Since the report size of the proposed PVFS and the proposed protocol are the same, the transmission failure is not based on the size of the report to be transmitted. To increase the reliability and realism of the experiment, experiments were conducted based on the maximum transmittable distance table according to the node output level. In the experiment of the proposed method, the list size is 70% of the maximum transfer range and the number of households is set to not exceed 50 generations. Rather than focusing on reducing routing costs, we focused on speeding up the routing updates. If the cost is less than the existing routing, the method is applied to the network immediately without entering the next generation so the routing can be formed more promptly.

4.3. Experimental result

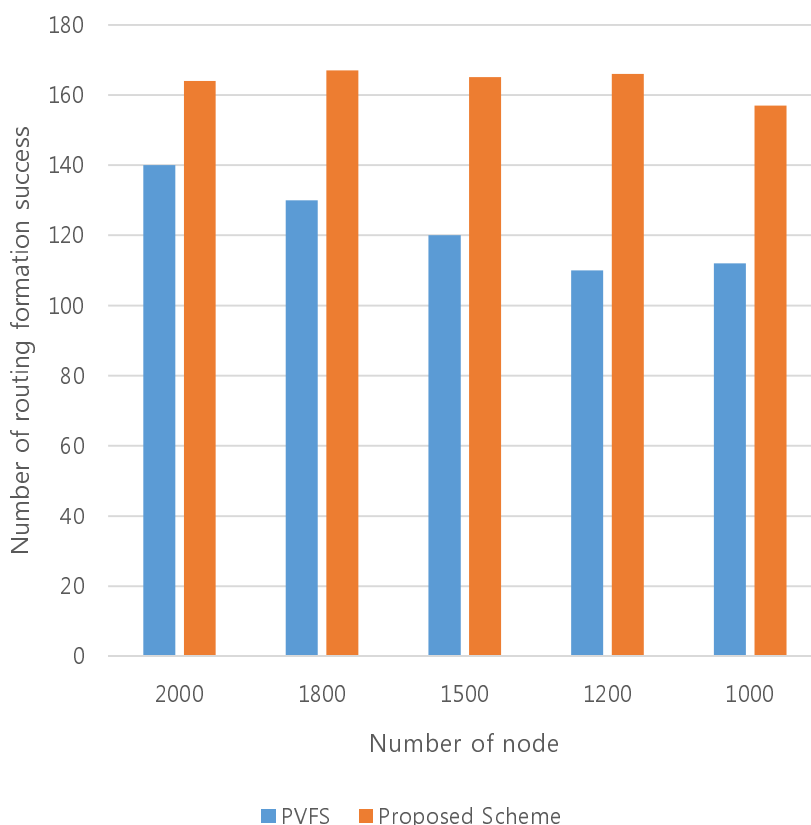


Figure 5. Number of routing formation successes.

Figure 5 shows the number of routing formation successes according to the number of nodes placed in the field. Overall, the difference in successes between the existing routing scheme and the proposed routing scheme is approximately 25%. It is possible to confirm the consistency of the routing formation success rate, irrespective of the number of nodes. In the conventional scheme, regardless of the number of nodes, the node with the smallest ID in one cluster is selected as CH, and thus does not affect the routing formation. In the case of the proposed scheme, even if the number of nodes is reduced, the performance is constant because the number of nodes is still high.

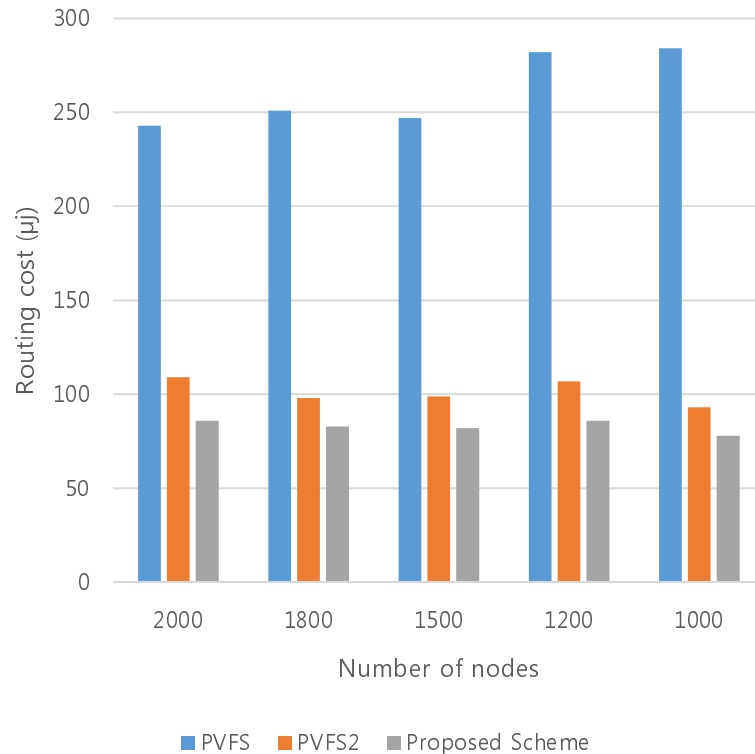


Figure 6. Routing costs depending on the number of nodes on the field.

Figure 6 shows a comparison of the routing costs consumed by each protocol depending on the number of nodes on the field. Comparisons were made only among successful routing cases. In the conventional scheme, energy consumption is independent of the number of nodes, and the number of nodes and routing cost are not related. In the case of the proposed method, as the number of generations of GA increases, the energy consumption gradually increases. However, in the experiment, the creation of the child chromosome for finding a better route path occurs immediately when the energy consumption is lower than that of the conventional method. In the energy consumption comparison experiment, the comparison with existing PVFS without the output level difference application was performed. Moderating the output of the node shows twice as much energy savings. In existing and proposed schemes with different output levels, the overall cost difference is approximately 20% due to the location correction of the event detection CH and the optimization of the routing cost optimization.

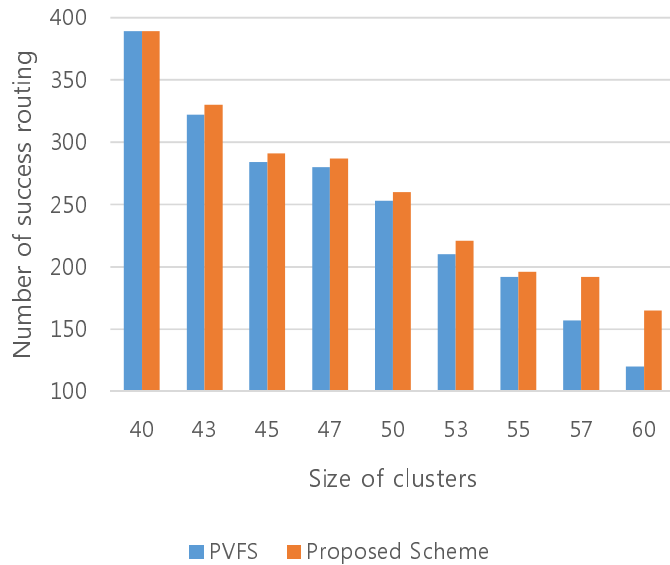


Figure 7. Number of routing successes according to the cluster size.

Figure 7 shows the number of routing successes according to the cluster size. When the cluster size is relatively small, there is not a large difference in the number of routing successes between the two schemes. However, as the size of the cluster increases, the number of routing successes increases. The number of successes according to the size of the existing scheme was greatly reduced because a node that does not reach the transmission range frequently occurs due to the randomly set CH. Alternatively, the proposed method does not have a large drop in selectivity of the forwarding node in the entire cluster, not just the CH. Additionally, because the algorithm is focused on immediate routing rather than routing costs, it is expected to be less if the system is set for cost reduction purposes.

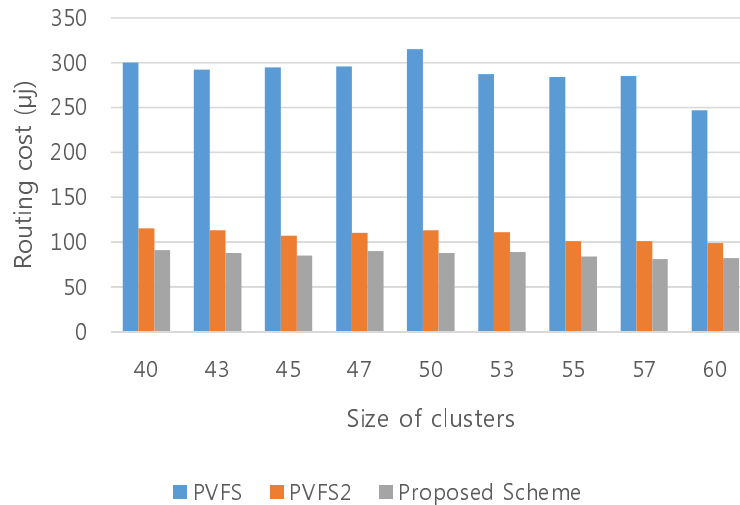


Figure 8. Difference in routing costs according to cluster size.

Figure 8 shows the difference in routing costs according to cluster size. This is similar to the figure of the routing cost according to the number of nodes, but the cluster size affects the entire routing success failure and can be interpreted as independent of the routing cost. Routing optimization of each node was completed without exceeding 10 generations. Routing of all CHs took less than 1 minute based on the field where more than 2000 nodes were dispatched.

5. CONCLUSIONS

Conventional cluster-based protocols are characterized by routing proceeding through CHs only, and the dependency on the usability of the CHs is so high that if the CHs are compromised, the security and energy efficiency of the entire routing using the nodes are damaged. Additionally, the higher energy usage of CH is expected when it is close to the BS. In this case, faster battery depletion occurs within those CHs, and the network users have to reroute frequently. In addition, there is also unnecessary energy consumption, which does not reflect the distance between nodes, but makes all transmissions the maximum output level. To solve these problems, the proposed scheme selects CH based on regional suitability, and selects proper delivery nodes considering every node in the network field. Cost optimization of routing was performed using the GA to optimize hop count and communication distance. With a proposed scheme, optimization of the routing cost can be confirmed in the situation where the size of the cluster changes or the number of nodes in the field changes.

ACKNOWLEDGEMENTS

This research was supported by the MISP (Ministry of Science, ICT & Future Planning), Korea, under the National Program for Excellence in SW) (2015-0-00914) supervised by the IITP (Institute for Information & communications Technology Promotion) (2015-0-00914).

REFERENCES

- [1] Al-Karaki, Jamal N., and Ahmed E. Kamal. "Routing techniques in wireless sensor networks: a survey." *IEEE wireless communications* 11.6 (2004): 6-28.
- [2] Perrig, Adrian, John Stankovic, and David Wagner. "Security in wireless sensor networks." *Communications of the ACM* 47.6.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] Yu, Zhen, and Yong Guan. "A dynamic en-route scheme for filtering false data injection in wireless sensor networks." *Proceedings of the 3rd international conference on Embedded networked sensor systems*. ACM, 2005.
- [7] Geetu, Sonia Juneja. "Performance analysis of SPIN and LEACH routing protocol in WSN." *International Journal of Computational Engineering Research* 2.5 (2012): 1179-1185.

- [8] Royer, Elizabeth M., and Chai-Keong Toh. "A review of current routing protocols for ad hoc mobile wireless networks." *IEEE personal communications* 6.2 (1999): 46-55
- [9] <https://www.scribd.com/document/91069327/Micaz-Datasheet-t>
- [10] D. E. Golberg, "Genetic algorithms in search, optimization, and machine learning," Addison Wesley, vol. 1989, 1989.
- [11] Deb, Kalyanmoy, et al. "A fast and elitist multiobjective genetic algorithm: NSGA-II." *IEEE transactions on evolutionary computation* 6.2 (2002): 182-197.
- [12] Whitley, Darrell. "A genetic algorithm tutorial." *Statistics and computing* 4.2 (1994): 65-85.
- [13] Benkic, Karl, et al. "Using RSSI value for distance estimation in wireless sensor networks based on ZigBee." *Systems, signals and image processing, 2008. IWSSIP 2008. 15th international conference on. IEEE, 2008.*
- [14] Buchli, Bernhard, Felix Sutton, and Jan Beutel. "GPS-equipped wireless sensor network node for high-accuracy positioning applications." *European Conference on Wireless Sensor Networks. Springer, Berlin, Heidelberg, 2012.*
- [15] Seon-Phil Kim, Namgi Kim. "Measurement and analysis of the data transmission range using the transmission power of MICAz." *Proceedings of KIIT Summer Conference. 2010*
- [16] Lin, Shan, et al. "ATPC: adaptive transmission power control for wireless sensor networks." *Proceedings of the 4th international conference on Embedded networked sensor systems. ACM, 2006*
- [17] Kyoung-Min Park, et al. "Path-Loss Measurements in Indoor Floor Environment at 2.4 GHz " *The Journal of Korean Institute of Communications and Information Sciences* 42.8 (2017): 1662-1667.
- [18] Mallinson, Michael, Patrick Drane, and Sajid Hussain. "Discrete radio power level consumption model in wireless sensor networks." *Mobile Adhoc and Sensor Systems, 2007. MASS 2007. IEEE International Conference on. IEEE, 2007*
- [19] Kubisch, Martin, et al. "Distributed algorithms for transmission power control in wireless sensor networks." *Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE. Vol. 1. IEEE, 2003*
- [20] Lin, Shan, et al. "ATPC: adaptive transmission power control for wireless sensor networks." *Proceedings of the 4th international conference on Embedded networked sensor systems. ACM, 2006*
- [21] Ogundile, Olayinka O., and Attahiru S. Alfa. "A survey on an energy-efficient and energy-balanced routing protocol for wireless sensor networks." *Sensors* 17.5 (2017): 1084.
- [22] Yadav, Anju, Vipin Pal, and Ketan Jha. "A Chaotic Genetic Algorithm for Wireless Sensor Networks." *Proceedings of First International Conference on Smart System, Innovations and Computing. Springer, Singapore, 2018.*
- [23] Srideviponmalar, P., V. Jawahar Senthil Kumar, and R. Harikrishnan. "Hybrid Genetic Algorithm–Differential Evolution Approach for Localization in WSN." *Intelligent Engineering Informatics. Springer, Singapore, 2018. 263-271.*

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.

