Comparative evaluation of information dissemination methods for effective and efficient information sharing in wireless sensor networks

Tomoyuki Tanabe, Naoki Wakamiya and Masayuki Murata
Graduate School of Information Science
and Technology, Osaka University
1-5 Yamadaoka, Suita,
Osaka 565-0871, Japan
Email: {t-tanabe, wakamiya, murata}@ist.osaka-u.ac.jp

***Abstract*—Wireless sensor networks (WSNs) consist of small nodes with sensing, computation, and wireless communication capabilities. Since overlaying multiple service-oriented networks on a WSN wastes bandwidth and energy, we consider all-to-all type information sharing in this paper. Although there have been various proposals of mechanisms to disseminate information among nodes in an efficient and effective manner, their performance has evaluated under specific conditions. In this paper, we conducted comprehensive evaluation of information dissemination methods to clarify their comparative characteristics and the range of application. Through simulation experiments, we showed that tree-based methods could achieve high delivery ratio in a small network, whereas a flooding-based or gossiping-based method was more effective in large-scale and low-density networks.**

I. INTRODUCTION

Wireless sensor networks (WSNs) consist of a large number of small nodes with sensing, computation, and wireless communication capabilities. Since it is wasteful and redundant to deploy multiple and independent application-oriented WSNs consisting of dedicated nodes and devices in the same region, a multi-purpose WSN has been attracting researchers in recent years [1]. In a multi-purpose WSN, application or service-oriented networks are overlaid and they share nodes and devices. Each overlay network consists of nodes contributing to the application in the current deployment strategy and they exchange messages with each other in an overlay network to provide users with desired functions or services. Limiting message exchanges among nodes belonging to the same overlay would help in saving energy and bandwidth to some extent. However, as the number of applications increases, concurrent multiple overlays dissipate energy and bandwidth. In such a scenario of a large number of concurrent applications, it is more adaptive, flexible, and energy and bandwidth efficient to share all information among all nodes through all-to-all communication realized by information dissemination methods.

There have been many proposals for efficient information dissemination in WSNs [2] and they have advantages and disadvantages. For example, a flooding-based method is the most primitive method which is simple and easy to implement. However, their greedy information forwarding by broadcasting causes duplicated information reception, bandwidth starvation, and low delivery ratio, especially in a high-density network. A publish/subscribe-based method such as SPIN [3] was proposed to avoid redundant information transmission by introducing a handshaking procedure prior to information forwarding. A node with information to disseminate first checks whether any neighbor nodes have not received the information. Then it sends the information to neighbor nodes if it receives a request for information transmission. The handshaking helps in reducing the number of information transmissions and receptions, but it would consume more energy and bandwidth and take longer time than a simple flooding method in disseminating new informatio.

As discussed above, characteristics of information dissemination methods differ from each other and their performance depends on several conditions such as the size of region and the node density. Therefore, we need to carefully select a method fulfilling requirements of desired functions or services under the expected operational condition. However, the performance of methods is evaluated under a specific condition in preceding literatures and we cannot directly compare them.

In this paper, we conduct comprehensive evaluation of information dissemination methods to clarify their comparative characteristics and the range of application. We first classify existing information dissemination methods into six, i.e. flooding, gossiping, publish/subscribe, ring, tree, and cluster. Then we evaluate their model methods from viewpoints of the scalability by changing the size of observation region and the node density. As performance measures we used the ratio of receiving nodes and the ratio of active time to evaluate the efficiency of information dissemination.

The remainder of this paper is organized as follows. First, in section II, we briefly describe six categories of information dissemination methods. Next, in section III, we describe simulation settings and measures. Then, in section IV, we present simulation results and discuss their comparative performance. Finally, in section V, we provide concluding remarks and future work.

II. CLASSIFICATION OF INFORMATION DISSEMINATION METHODS

In this section, we describe six categories of information dissemination methods. Hereafter, information to be shared among nodes, e.g., sensing data, is called *data*, a node which generates data is called *source node*, and a message carrying data is called *data message*. Moreover, a message other than a data message is called *control message*. We assume that a data message does not contain any control information.

Flooding-based Method

With a flooding-based method, a source node broadcasts a data message to all of its neighbor nodes. A neighbor node which receives the data message for the first time forwards the data message to all of its neighbor nodes by broadcasting. Otherwise, it silently discards the data message. By repeating the forwarding procedure, the data message is eventually received by all nodes in the network under ideal condition. To accomplish all-to-all communication, all nodes become source nodes and initiate flooding.

Gossiping-based Method [4], [5], [6]

A gossiping-based method is similar to a flooding-based method, but message forwarding is done in a stochastic manner. With a gossiping-based method, a source node broadcasts a data message to all of its neighboring nodes. A neighbor node which receives the data message for the first time forwards the data message with probability p $(0 < p < 1)$ to all of its neighbor nodes by broadcasting. Otherwise, it silently discards the data message. As far as the forwarding probability p is sufficiently large, whose critical value can be given by the percolation theory [7], the data reachability can stochastically be guaranteed. As in the flooding-based method, all nodes become source nodes and initiate data dissemination for all-to-all communication.

Publish/Subscribe-based Method [3], [8]

Both of the above methods adopt broadcasting in forwarding a data message. As such, there is a possibility that a node receives the same data message several times especially in a densely connected network. To avoid the redundant message reception, a publish/subscribe-based method introduces a handshaking procedure before data message transmission. When a source node has new data to disseminate or a node receives new data, they first broadcast a small message, called *metadata*, which contains the information about the data to send, so that neighbor nodes can judge whether they need to receive the data or not. If a neighbor node has not received the data, it sends a request message to the sender of the metadata. Then, the sender sends the data message to the requesting node. There are variants of publish/subscribe-base methods, which differ in the way that a node sends control and data messages, such as SPIN-PP and SPIN-EC [8], we base our discussion on a method which adopts broadcasting in control and data message transmission. All-to-all communication can be accomplished by initiating data dissemination at all nodes.

Ring-based Method [9], [10]

Differently from the above three methods, the following three methods relies on the topological structure of a network for efficient data dissemination. With a ring-based methods, all nodes in a network form a ring over the physical network topology. Independently of the actual number of physical neighbor nodes within the range of radio communication, each node has only two neighbors on a ring. First a source node sends a data message to two adjacent nodes in unicast communication. Next, a node receiving the data message forwards it to the neighbor node on the other side if it is the first reception. Two data messages traverse the ring both clockwise and counterclockwise respectively, and they eventually meet at the node locating at the opposite side of the source node on the ring. At this time, data dissemination is considered finished. We should note here that a mechanism and corresponding control overhead of ring construction are out of scope of the paper. In the ring-based method, all nodes start broadcasting its data message for all-to-all communication.

Tree-based Method [11]

With a tree-based method, a single tree topology which consists of all nodes in a network is constructed. A data message first goes up to a root of the tree and then it is distributed to all nodes. A source node first sends a data message to its parent node in the tree. If it is the first time that a parent node receives the data, it forwards the data message to its parent node. By repeating the process, the data message finally arrives at the root node. Then, the root node broadcasts the data message to all child nodes. They forward the data message by broadcasting if they have not done. Eventually, all leaf nodes located at the bottom of the tree receive the data message and data dissemination is completed at this time. As for the ring-based method, we do not take into account tree construction overhead in the evaluations. In the case of a tree-based method, not only superimposing one-to-all data dissemination but data aggregation can be used to accomplish all-to-all communication. Details of data aggregation will be given in section III-C.

Cluster-based Method [12], [13], [14]

To save energy consumption in data dissemination and gathering, many researchers consider cluster-based methods are the most promising [15]. Nodes are grouped into clusters with a cluster-based method in accordance with their proximity and one node is appointed as a *cluster head* in each cluster. A cluster head is responsible for data dissemination and gathering within its cluster and data exchange among clusters. The other nodes in the cluster are called *cluster member*. Although there are a variety of clustering methods proposed in literatures, in this paper we consider a method explained below as a typical and representative method.

The method consists of two phases. First in the clustering phase, a certain number of nodes in a network elect themselves as cluster heads by, for example, an algorithm used in LEACH [12], and broadcast an advertisement message. Other node receiving the advertisement message becomes a cluster member of the sender. A node which receives two or more advertisement messages is called border node. It becomes a cluster member of a cluster head with the highest signal strength and participates in inter-cluster message transmission in the data transmission phase. Then, a cluster member sends a join message to the cluster head.

Once clusters are organized, next in the data transmission phase, a cluster member, i.e. a source node, sends a data message to its cluster head. Next, a cluster head broadcasts a data message, which contains both of data received from cluster members and its own data, to all cluster members. A border node in the cluster then forwards the data message to the other cluster heads from which it received advertisement messages. When there are two or more border nodes in a cluster, one of them is appointed as a forwarder by a cluster head. If it is the first time that a neighbor cluster head receives the data message, it aggregates all data it has and broadcasts an aggregated data message to cluster members. Consequently, data is disseminated to all nodes in the network.

III. EVALUATION SETTINGS AND MEASURES

In this section, we describe simulation settings and evaluation measures. In the experiments, we use ns-2.34.

A. Arrangement of Nodes

To consider various operating conditions, we use the following two scenarios. In both scenarios, nodes communicate with each other by the IEEE 802.11 MAC protocol with RTS/CTS. The communication range is set at 100 [m] and the transmission speed is 1 [Mbps].

The first scenario is to change the size of observation region. We call an area of 40 [m] × 40 [m] “block”. One node is placed at a random location of each block. The observation region is constituted by arranging blocks in a square. The size of observation region is changed by the number of blocks, from 8 × 8 blocks, i.e. 64 nodes in the region of 320 [m] × 320 [m] to 30 × 30 blocks, i.e. 900 nodes in the region of 1200 [m] × 1200 [m]. The node density is 0.000625 [node/m^2].

In the second scenario, the node density is changed while the size of the observation region is kept constant at 480 [m] × 480 [m]. The observation area is divided into blocks and each block has a randomly placed node. We change the number of blocks and the resulting node density from 8 × 8 blocks (64 nodes) and 0.00028 [node/m^2] to 30 × 30 blocks (900 nodes) and 0.0039 [node/m^2], respectively.

B. Data Generation Model

Performance of information dissemination methods heavily depends on the way that data are generated at nodes. If all nodes begin sending a data message at the same time, information dissemination suffers from collisions and congestions and all-to-all communication easily fails. In this paper we consider asynchronous data generation, which is more realistic than the synchronous data generation.

With the asynchronous data generation model, each node generates new data at random time from 0 to 1 [s] of simulation time. In the case of a tree-based method, leaf nodes begin to send a data message once data is generated and other nodes wait for data message reception from child nodes. In the case of a cluster-based method, cluster members begin to send a data message on generation and cluster heads wait for data message reception from cluster members. In the other methods, all nodes begin to send a data message when it is generated. The size of a data message is set at 1 [Kbyte].

C. Data Aggregation

For efficient information dissemination, a parent node in a tree-based method and a cluster head in a cluster-based method waits for reception of data messages from child nodes or cluster members, aggregate them with its own data, and then forwards the aggregated data to a further parent or another cluster head by a border node, respectively. However, it is not guaranteed that all child nodes or cluster members successfully transmit their data messages to a parent node or a cluster head due to collision and congestion. Therefore, we introduce a timer to limit waiting time.

In the case of a tree-based method, a node located at b hops from the root node in a tree whose height is a waits for data messages for $a-b$ [s] from when it received a data message from a child node for the first time. Data received during this period are aggregated and forwarded to a further parent node, while data messages arrived after timer expiration are immediately discarded. In the case of a cluster-based method, a cluster head waits for data messages for 1 [s] from when it received a data message from a cluster member for the first time. A cluster head aggregates data received in the 1-second waiting time and broadcasts the aggregated data to cluster members. When it receives a data message after that, it broadcasts the message immediately.

D. Evaluation Measures

To evaluate performance of information dissemination methods in all-to-all communication, we use the ratio of receiving nodes and the ratio of active time.

We first define $R(t,i,j)$ by Eq. (1). It represents whether node j has received data generated by node i by time t [s].

$$R(t,i,j)=\begin{cases}1, & \text{if i} \neq \text{j and received}\\ 0, & \text{if } i=j \text{ or unreceived}\end{cases} \tag{1}$$

Furthermore, we define $T(i)$, which is time for dissemination of data generated by node i to be completed. Then, the ratio of receiving nodes is derived as follows.

$$D=\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}R(T(i),i,j) \tag{2}$$

where n is the number of nodes in a network.

The ratio of active time is defined as follows,

$$W=\frac{1}{n^2}\sum_{i=1}^{n}\frac{\sum_{j=1}^{n}A(i,j)}{T(i)} \tag{3}$$

where $A(i, j)$ represents the sum of time spent by node j from beginning of carrier sense to completion or cancellation of message transmission, from reception of a RTS message to completion or cancellation of ACK transmission in unicasting, and from beginning to end of reception of a message in broadcasting, in dissemination of data of node i.

IV. SIMULATION RESULTS

In this section, we show results of simulation experiments. Throughout the experiments, the forwarding probability p is set at 0.5 for a gossiping-based method and the size of metadata is set at 16 [Byte] for a publish/subscribe-based method. Regarding topology-dependent methods, information dissemination is initiated after topology construction and there is no topology change during a simulation run. Therefore, only control messages exchanged in handshaking with a publish/subscribe-based method are taken into account in simulation experiments. In the following, results averaged over 50 simulation runs are shown for each method and setting.

A. Changing the size of observation region

As shown in Fig. 1(a), for a network with less than 400 nodes, a tree-based method accomplishes the highest ratio of receiving nodes. A reason why the performance drastically deteriorates with more than 300 nodes is that there occur collisions among data messages going up toward the root node and those going down toward leaf nodes. Although a node discards data messages received after the waiting time to avoid such collisions, the waiting time becomes insufficient for a tree of high height. A ring-based method also suffers from the increase in the number of nodes. Since a node forwards a data message right after reception, it experiences more collisions than a tree-based method and gives up transmission for backoff counter limitation. A reason why a cluster-based method results in the lowest performance is in inter-cluster message transmission. Message broadcasting from a border node to neighbor cluster heads and from a cluster head to border nodes often collide with each other. As a result, data are only shared among nodes belonging to the same cluster.

Differently from the topology-dependent methods, we can observe the increase in the ratio of receiving nodes with a flooding-based, gossiping-based, and publish/subscribe-based methods. The reason is that they adopt broadcasting in data message forwarding. Broadcasting does not negotiate channel usage by RTS/CTS with neighbor nodes, confirm reception of data message, or MAC-layer retransmission as unicasting does. As such, the ratio of receiving nodes is lower than the tree-based method and the ring-based method for cases with the small number of nodes. However, the diffusibility is stronger for its greedy message forwarding especially in a network with a large number of nodes. Only if one of neighbor nodes of a sender can successfully receive a data message, message dissemination continues. The other neighbors that fail in receiving a data message directly from a sender are likely to receive the message afterward from the successful receiver or other neighbors after several steps of message forwarding. Moreover, since collisions are likely to occur around a source, a data message that successfully reaches a node further from a source node can diffuse itself in the area.

A gossiping-based method achieves the lower ratio of receiving node than a flooding-based method in most of the cases. It is because a node does not always forward a data message with a gossiping-based method. However, such modest forwarding results in the higher performance at the rightmost point and the gradual decrease in the ratio of receiving nodes. A reason that a publish/subscribe-based method cannot achieve the high performance is the handshaking process. Since all data generated at nodes are new to all nodes in a network, a large number of control messages are exchanged and they disturb data message transmission.

Figure 1(b) shows that the ratio of active time is not affected by the number of nodes independently from adopted methods. A ring-based method results in the highest ratio of active time, because all message transmissions are in unicast which takes longer time than broadcasting. In the case of a publish/subscribe-based method, handshaking lengthen a process of message transmission. As a result, the ratio of active time becomes large.

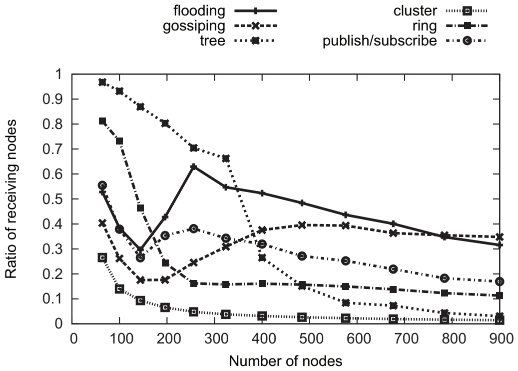
In conclusion, when the size of observation region is smaller than 800 [m] × 800 [m] (400 nodes), a tree-based method brings the highest ratio of receiving nodes with the low ratio of activation time. For a larger region, a flooding-based method or a gossiping-based method is more efficient.

B. Changing the node density

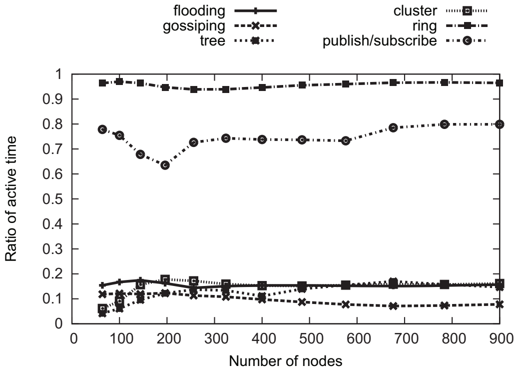
As shown in Fig. 2(a), a tree-based method accomplished the highest ratio of receiving nodes independently from the node density. It is because the tree height remains the same against density changes, while the higher tree causes performance deterioration in Fig. 1(a). A ring-based method achieves the slightly higher ratio of receiving nodes among the remaining methods for unicast-based message forwarding. The other methods suffer very much from the increase in the node density. Broadcasting is apparently an ineffective mean of message transmission in a dense network.

Regarding the ratio of active time in Fig. 2(b), a tree-based method has the lowest ratio in a sparse network but the performance gets worse as the node density increases. It is because that a node has to wait longer for a wireless channel to be available to unicast a data message to a parent node in a dense network. With the similar reason, the ratio of active time also increases with the ring-based method. On the contrary, the ratio decreases with the publish/subscribe-based method. As the node density increases, a node has less chance to receive a data message. Then, it does not need to forward a data message. As a result, time consumed in handshaking and message forwarding decreases.

In conclusion, a tree-based scheme is the most preferable in the high density network, although the ratio of receiving nodes is not high enough.

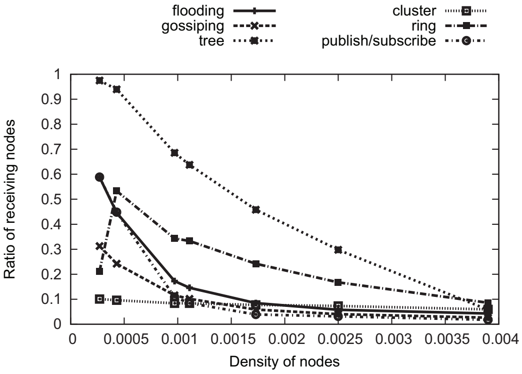


(a) The ratio of receiving nodes

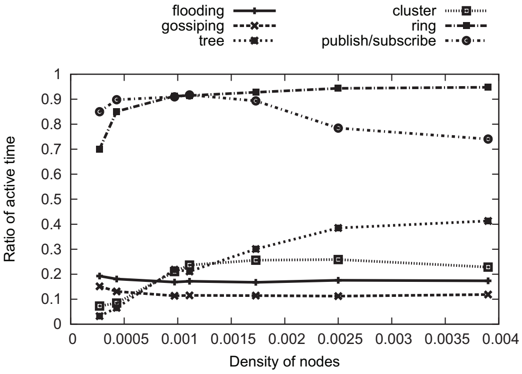


(b) The ratio of active time

Fig. 1. Comparative results by changing size of observation region



(a) The ratio of receiving nodes



(b) The ratio of active time

Fig. 2. Comparative results by changing the node density

C. Influence of node location

Finally, we evaluate the influence of node location in information dissemination. For this purpose, we additionally define two measures, d_i and r_i.

$$d_i = \frac{1}{n-1}\sum_{j=1}^{n} R(T(i), i, j) \tag{4}$$

$$r_i = \frac{1}{n-1}\sum_{j=1}^{n} R(T(j), j, i) \tag{5}$$

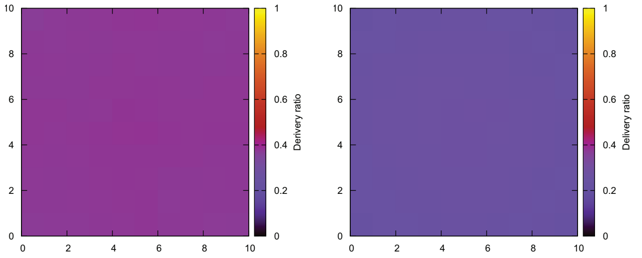
d_i, called *delivery ratio*, indicates the ratio that data generated at node i is received by the other nodes in a network. On the other hand, r_i, called *reception ratio*, indicates the ratio that node i receive data generated at the other nodes in a network. Figures 3 and 4 illustrates distribution maps of the delivery ratio and the reception ratio for the case of 100-node networks, respectively. The color of a block shows the average value over 10 simulation runs.

As shown in Fig. 3, the location of node does not affect the delivery ratio very much independently of methods. In general, nodes located at the center have more neighbor nodes than those located near the edge. The larger number of neighbors contributes to higher survivability of data for having more paths to other nodes. However, with the asynchronous generation model, message transmission would experience more collisions and congestions and can fail. As a result, we see the even distribution of delivery ratio.

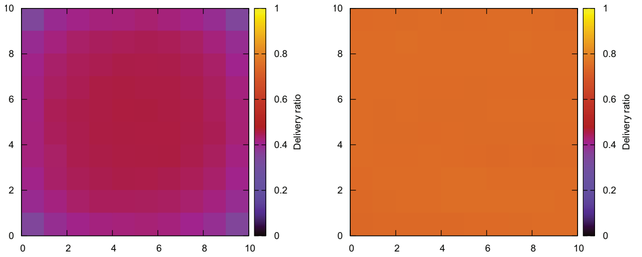
On the contrary, nodes located at the center of a network can receive information from more nodes than those located near the edge with the flooding-based, gossiping-based, and publish/subscribe-based methods in Fig. 4. It is because nodes located at the center can receive data only if any one of neighbors can successfully receive and transmit a data message. In the case of topology-dependent methods, they can accomplish rather even delivery ratio, which is higher than the highest ratio with the other tree methods.

V. CONCLUSION

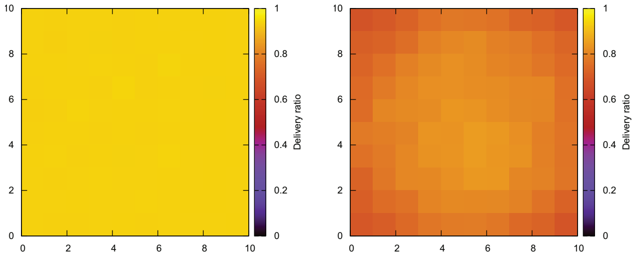
In this paper, we conducted comparative evaluations of well-known information dissemination methods for all-to-all communication in a multipurpose WSN. As a result, we revealed that a tree-based method well fits to the small observation region independently from the node density and a flooding-based method can achieve the higher ratio of receiving nodes in the large observation region. However, none of methods could accomplish the perfect information dissemination. We plan to consider a method which guarantees the reliable information dissemination for all-to-all communication.



(a) Flooding-based method (b) Gossiping-based method

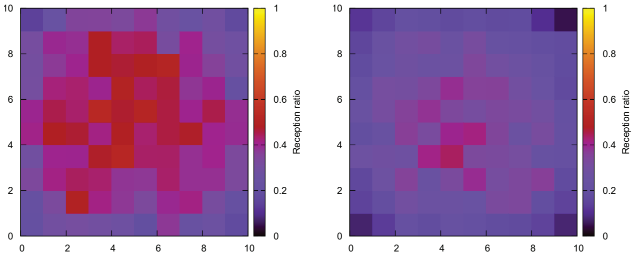


(c) Publish/Subscribe-based method (d) Ring-based method

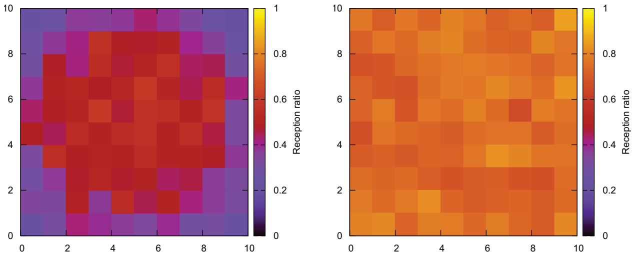


(e) Tree-based method (f) Cluster-based method

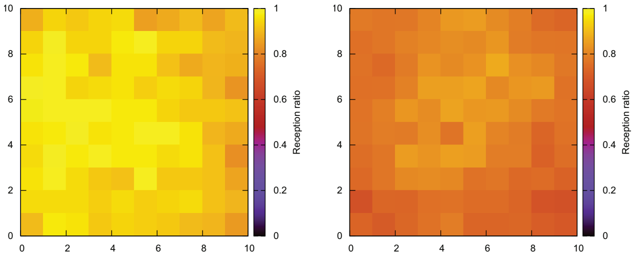
Fig. 3. Influence of node location on delivery ratio



(a) Flooding-based method (b) Gossiping-based method



(c) Publish/Subscribe-based method (d) Ring-based method



(e) Tree-based method (f) Cluster-based method

Fig. 4. Influence of node location on reception ratio

ACKNOWLEDGMENT

This research was supported in part by Grand-in-Aid for Scientific Research (B) 22300023 of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

REFERENCES

[1] J. Steffan, L. Fiege, M. Cilia, and A. Buchmann, "Towards multi-purpose wireless sensor networks," in *Proceedings of Systems Communications*, pp. 336–341, Aug. 2005.

[2] N. Gautam, S. Sofat, and R. Vig, "Data dissemination in wireless sensor networks: A state-of-the art survey," *International Journal of Computer Networks*, vol. 4, no. 1, pp. 22–34, 2012.

[3] W. Heinzelman, J. Kulik, and H. Balakrishnan, "Adaptive protocols for information dissemination in wireless sensor networks," in *Proceedings of ACM annual ACM/IEEE international conference on Mobile computing and networking*, pp. 174–185, June 1999.

[4] S. Hedetniemi, S. Hedetniemi, and A. Liestman, "A survey of gossiping and broadcasting in communication networks," *IEEE Networks*, vol. 18, no. 4, pp. 319–349, 1988.

[5] Z. Haas, J. Halpern, and L. Li, "Gossip-based ad hoc routing," *IEEE/ACM Transactions on Networking*, vol. 14, pp. 491–450, June 2006.

[6] A. Dimakis, A. Sarwate, and M. Wainwright, "Geographic gossip: efficient aggregation for sensor networks," in *Proceedings of IEEE International Conference on Information Processing in Sensor Networks*, pp. 69–76, Apr. 2006.

[7] D. Stauffer and A. Aharony, *Introduction to percolation theory*. 1991.

[8] J. Kulik, W. Heinzelman, and H. Balakrishnan, "Negotiation-based protocols for disseminating information in wireless sensor networks," *Wireless Networks*, vol. 8, pp. 169–185, Mar. 2002.

[9] S. Huang, R. Jan, and W. Yang, "RICA: A ring-based information collection architecture in wireless sensor networks," *Sensor Networks*, vol. 1, pp. 190–199, Jan. 2006.

[10] R. Nowak, "Distributed EM algorithms for density estimation and clustering in sensor networks," *IEEE Transactions on Signal Processing*, vol. 51, pp. 2245–2253, Aug. 2003.

[11] T. Banerjee, K. Chowdhury, and D. Agrawal, "Tree based data aggregation in sensor networks using polynomial regression," in *Proceedings of IEEE International Conference on Information Fusion*, pp. 25–29, July 2005.

[12] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proceedings of IEEE Annual Hawaii International Conference on System Sciences*, pp. 1–10, Jan. 2000.

[13] A. Manjeshwar and D. Agrawal, "TEEN: A routing protocol for enhanced efficiency in wireless sensor networks," in *Proceedings of IEEE International Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing*, pp. 2009–2015, Apr. 2001.

[14] O. Younis and S. Fahmy, "HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Transactions on Mobile Computing*, vol. 3, pp. 366–379, Oct. 2004.

[15] A. Abbasi and M. Younis, "A survey on clustering algorithms for wireless sensor networks," *Computer communications*, vol. 30, pp. 2826–2841, June 2007.