PAPER

On Characteristics of Multi-hop Communication in Large-scale Clustered Sensor Networks

Yuichi KIRI[†a], ***Nonmember***, **Masashi SUGANO**[††b], ***Member***, ***and* Masayuki MURATA**[†c], ***Fellow***

SUMMARY After the popularization of sensor networks, the size of the monitoring region and the number of sensor nodes will grow to an enormous scale. In such large-scale sensor networks, multi-hop communications between sensor nodes will be necessary to cover the whole monitoring region. Moreover, sensor nodes should be grouped into clusters to enhance scalability and robustness. Therefore, we believe that multi-hop communication between clusters is preferable for large-scale sensor networks. To clarify the fundamental characteristics of this form of communication, we analytically derive the network's power consumption and compare it with other routing methods using TDMA communication and an interference-free transmission schedule. The results show multi-hop communication between clusters is preferable in large-scale sensor networks because it can alleviate heavy relaying loads near the sink node and it has a shorter data collection time compared with simple multi-hop communications without clusters.

Knowing how much performance degradation arises when interference is unavoidable is essential for multi-hop communications in clustered sensor networks. Therefore, we compare interference-free TDMA communication with CSMA/CA communication which can cause interference in clustered sensor networks. Consequently, we show that although the data collection time is about 3.7 times longer when using CSMA/CA, the power consumption can be suppressed to 12%.
key words: *sensor network, clustering, multi-hop communication, simulation, interference*

1. Introduction

Sensor networks have uses in disaster prediction, security, environmental monitoring, and traffic control. A wide variety of network sizes are used in these applications. In environmental monitoring, for example, hundreds or thousands of sensor nodes are deployed in a large monitoring region. In such large-scale sensor networks, scalability and robustness are very important. In addition, sensor nodes are highly power-constrained, and they must work at very low power to prolong the lifetime of the sensor network. Clustering, a method of grouping sensor nodes, can meet these requirements, and it has been the focus of much research on sensor networks [1]–[7]. In addition, the communication ranges of sensor nodes are generally short. Thus, to collect data from a large monitoring region, multi-hop communication is necessary.

Multi-hop communication between clusters is preferable for data collection in large-scale sensor networks, and we expect that this form of communication will become widespread. On the other hand, further evaluations from a broad perspective are required to promote multi-hop communication between clusters. For example, from the viewpoint of power consumption, it is crucial to consider the distribution of power consumed in each area rather than the total power consumption over a monitoring region. Also, to be able to lay a sensor network with multi-hop communication between clusters, we need to understand the effect that the distribution has on the data collection rate and how long it takes from the moment a sensor node senses the monitoring region until the data reaches the sink node. To the best of our knowledge, the above matters have not been sufficiently studied in previous researches.

The purposes of this research are to clarify the fundamental characteristics of multi-hop communications between clusters and to confirm the effectiveness and the domain of applicability. We analytically derived the power consumption of clustered multi-hop sensor networks in [8]. Based on an ideal situation where no interference occurs, this analysis states the relationship between cluster size and power consumption. Therefore, by comparing it with simulation results using TDMA based on an interference-free ideal schedule, we can provides useful characteristics of such a communication with the analysis as a unit, as well as correctness of our simulation programs. Moreover, we also show the advantages and disadvantages of such a communication form by comparing it with basic routing methods using the ideal TDMA. We assumed that such an ideal schedule has positional information available for every sensor nodes.

For a realistic evaluation, however, it is important to determine how much performance degradation occurs when the positional information on all sensor nodes is not available, that is, when interference is inevitable. Constructing an ideal interference-free trans-

Manuscript received June 30, 2006.
Manuscript revised August 21, 2006.
Final manuscript received October 10, 2006.
[†]The authors are with the Graduate School of Information Science and Technology, Osaka University, 1-5 Yamadaoka, Suita-shi, Osaka, 565-0871, Japan
[††]The author is with the School of Comprehensive Rehabilitation, Osaka Prefecture University, 3-7-30 Habikino, Habikino-shi, Osaka, 583-8555 Japan
a) E-mail: y-kiri@ist.osaka-u.ac.jp
b) E-mail: sugano@rehab.osakafu-u.ac.jp
c) E-mail: murata@ist.osaka-u.ac.jp

mission schedule is impractical in large-scale sensor networks. In fact, each sensor node must repeatedly broadcast its information in such a network, and the power consumed by these broadcasts surely decreases the network lifetime. Moreover, each sensor node requires high-capacity memories to store many sensor nodes' information but severe constraints exist regarding memory capacity of the sensor node. Therefore, the ideal schedule is unrealizable. In large-scale sensor networks, the number of sensor nodes within communication range is large. If the ideal schedule is not available in such a situation, a considerable amount of interference would occur and performance degradation would be unavoidable. We think that an understanding of how much performance is lost would be extremely useful for laying clustered multi-hop sensor networks.

The rest of the paper is organized as follows. In Section 2, we describe related studies on clustering and multi-hop communication between clusters. In Section 3, we explain our network model. In Section 4, we clarify the fundamental characteristics of sensor networks with multi-hop communication between clusters by comparing them with other routing methods through analysis and simulation experiments. In Section 5, we discuss the degree of performance degradation that occurs with CSMA/CA in comparison with an ideal transmission schedule with a TDMA mechanism. Section 6 concludes the paper.

2. Related work

In sensor networks, a low-energy adaptive clustering hierarchy (LEACH) has been proposed as a clustering method for reducing the power consumption of sensor networks [5]. In LEACH, each sensor node decides whether to become a cluster head based on a predetermined percentage. Clusters are constructed with sensor nodes adjacent to cluster heads. Communication from sensor nodes to cluster heads and communication from cluster-heads to a sink node is performed via a single hop.

Many clustering methods aimed at improving on LEACH have been suggested. In LEACH, the positions of clusters can be unbalanced, decreasing the network lifetime. Hybrid energy-efficient distributed clustering (HEED) [7] places clusters uniformly over a monitoring region. Furthermore, HEED balances power consumption between sensor nodes. The central controlling algorithm, which provides a regular cluster size, was also proposed [4]. The time complexity of this algorithm is $O(n^3)$, where n is the number of sensor nodes. Although LEACH uses single-hop communication within a cluster, hybrid indirect transmissions (HIT) [3] uses multi-hop communication within clusters to limit the interference range and to communicate in parallel with as many nodes as possible.

The transmission distance must be reduced to minimize the power consumption of sensor networks. From this standpoint, power-efficient gathering in sensor information systems (PEGASIS) [6] comprises a chain instead of a set of clusters. This chain connects the nearest neighboring sensor nodes, and the distance between sensor nodes is very short. Two-Phase Clustering (TPC) [2] also constructs chains within clusters. Although these studies evaluated various clustering methods by using several metrics such as power consumption or data collection time, most of them focused on small sensor networks in which 1-hop communication between an arbitrary sensor node and the sink node can always be achieved. Although multi-hop communication between clusters is employed in [2], its performance is not specifically mentioned.

Studies have also been done on multi-hop communication between clusters in sensor networks. In connectionless probabilistic (CoP) routing [9], the monitoring region is divided into square areas, and multi-hop communication takes place between cluster heads positioned at the vertices of these areas. Neander *et al.* evaluated a sensor network using multi-hop communication between cluster heads mainly through simulation experiments [10] that assumed the sink node can directly communicate with all sensor nodes. The unequal clustering size (UCS) [11] was designed to equalize power consumption among cluster heads. In UCS, a circular monitoring region is split into two concentric circles, called layers. Soro and Heinzelman determined that the size of the cluster in the interior layer should be reduced to equalize the power consumption. Shu *et al.* divided a monitoring region into multiple layers and derived optimal parameters, such as the cluster radius of each layer and the relay probabilities of cluster heads, to prolong the coverage-time [12]. These studies indicated that power consumption is a crucial metric in sensor networks that have severe power constraints. However, these evaluations were only for the power consumption of the whole sensor network. In sensor networks whose purpose is to collect data, these data are concentrated around the sink node and the residual power in this region determines the lifetime of a network. Evaluating the whole sensor network is not sufficient if the power consumption varies enormously from region to region; the per-region power consumption should be investigated instead. Additionally, we need to examine the impact of the power consumption on the data collection rate. Because some sensor network applications require prompt data gathering, we also regard the data collection time as an essential metric. However, few studies have evaluated it. We analytically derived the power consumption in clustered multi-hop sensor networks in [8], and here, we clarify the overall performance characteristics of clustered multi-hop sensor networks by using a wide variety of simulation experiments regarding the metrics stated above.

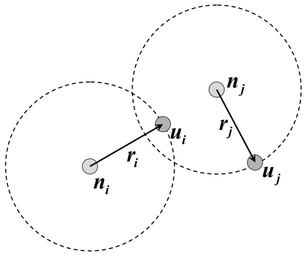


Fig. 1 Data interference.

3. Sensor network model

3.1 Network model

A model of the sensor network under consideration is as follows. The sink node is placed at the center of the region. Data fusion is not used to reduce the data volume. That is, the data generated by the sensor nodes is transmitted to the sink node without any modification or compression. The sensor nodes are placed randomly and uniformly throughout the area. We assume that both the sink node and the sensor nodes are stationary after deployment. All sensor nodes have the same initial power and communication capabilities. Moreover, they have the ability to control the transmission power depending on the distance between the sensor node and its next-hop node. We also assume that sensor nodes are synchronized with each other and they send and receive data in synchronization with fixed-length timeslots. We suppose the existence of a synchronization method, but an evaluation of its feasibility and overhead is outside the scope of our research because we are interested in the performance under ideal conditions. The same wireless channel is used in the entire network for the intra-cluster communication and another one is used between cluster heads.

When a sensor node receives multiple packets at the same time, the node cannot receive both packets correctly. We denote this situation as data interference, and we account for it by using the model presented in [3]. To illustrate data interference, consider four sensor nodes shown in Fig. 1. Sensor node n_i sends a packet to sensor node u_i located a distance r_i from n_i, and n_j sends a packet to u_j located at distance r_j from n_j. In this case, sensor node u_i receives packets from n_i and from n_j simultaneously and thus they interfere with each other. This condition is represented in (1). That is, the interference of data occurs when n_i and n_j, which satisfy (1) send data simultaneously. $d(u_i, n_j)$ in (1) corresponds to the distance between sensor node u_i and n_j.

$$d(u_i, n_j) < r_j \tag{1}$$

3.2 Power consumption model

The lifetime of a sensor networks depends on the operation times of individual sensor nodes. Therefore, a model that defines the amount of power consumed in each action of a sensor node influences the network lifetime to a great degree. We used the power consumption model represented in [5]. In this model, the power consumption of the radio in order to run the transmitter or receiver circuitry is equal to E_{elec}, and the power to run the transmit amplifier is equal to ϵ_{amp}. Using these parameters, the power needed to transmit k bits of data over a distance d is:

$$E_{T_x} = E_{elec}k + \epsilon_{amp}kd^2 \tag{2}$$

and the power needed to receive k bits of data is:

$$E_{R_x} = E_{elec}k \tag{3}$$

3.3 Routing

Using multi-hop communication between clusters, a sensor node must determine its next-hop node. For simplicity, only the distance to the next hop is used as a selection criterion in choosing a relay node. For example, the procedure for selecting a relay node within a cluster is as follows. In Fig. 2, n_s represents a sensor node that transmits a packet, r_{max} is the maximum communication range, n_i is the next-hop node of n_s, and CH_{n_s} represents the cluster-head of n_s. The next-hop node n_i must be closer to the cluster head than the sending node n_s, i.e.,

$$d(n_i, CH_{n_s}) < d(n_s, CH_{n_s}) \tag{4}$$

The power of the sending node needed to transmit a packet to its next-hop node is smaller than the power needed to transmit a packet to its cluster head, i.e.,

$$d(n_s, n_i) < d(n_s, CH_{n_s}) \tag{5}$$

At the same time, the distance between the sending node and next-hop node is bounded by

$$d(n_s, n_i) \leq r_{max} \tag{6}$$

because of constraints on the maximum communication range. n_s selects the nearest sensor node that satisfies the above three inequalities as its next-hop node. If n_i does not exist and $d(n_s, CH_{n_s}) \leq r_{max}$ is satisfied, the next-hop node of n_s is cluster-head CH_{n_s}. The same is true for selecting the next-hop cluster head in multi-hop communication between cluster heads.

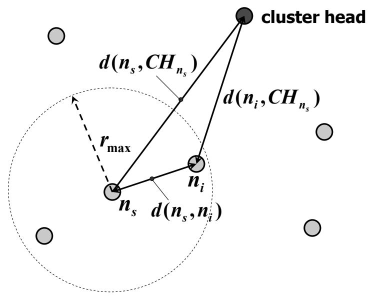


Fig. 2 Procedure for selecting relay node within a cluster.

4. Fundamental characteristics of multi-hop communication between clusters

4.1 Creation of transmission schedule without interference

In our interference model, the interference is only a function of the distance between sensor nodes. In this case, if the positional information for all sensor nodes is available, a transmission schedule that eliminates data interference can be constructed, and communications with a TDMA mechanism becomes possible. Here, we use the algorithm proposed in Culpepper *et al* [3] for autonomously constructing a transmission schedule. In our simulation experiments, each sensor node followed this schedule, and they transmitted packets to their next-hop sensor nodes in the assigned timeslots. The transmissions were free from interference, because we assumed that no transmission errors would occur and that packets would be received by the next-hop sensor nodes without errors. All packets arrived at the sink node at predetermined times.

4.2 Simulation setting

To clarify the basic characteristics of multi-hop communication between clusters, we used the following three routing methods as objects for comparison for LEACH+multi-hop (LEACH+MULTI) and HEED+multi-hop (HEED+MULTI).

- (DIRECT) Each sensor node communicates with the sink node directly.
- (LEACH) Once the clusters have been constructed, each sensor node transmits a packet to its cluster head, and the cluster heads send those packets and their own packet to the sink node directly.
- (MULTI) Each sensor node sends packets to the sink node in a multi-hop fashion without clustering.

Table 1 lists the parameters used in our simulations. We assumed that the sensor nodes perform sensing and generate packets simultaneously. After the sink node receives all the packets that can reach it, the sensor nodes perform sensing again. We denote this period as a cycle. In routing methods that construct clusters, the roles of a cluster head between sensor nodes have to be changed once in a period, in what is called a round, to prevent cluster heads from using up all their power. Although control signals, such as cluster-head advertisements, are sent and received every round and the sensor nodes consume power, we define the length of a round to be the same as the length of a cycle. That is, clustering is performed every cycle. Our simulation results are the average values obtained over 100 simulations.

Table 1 Parameter settings.

Parameter	Value
Radius of monitoring region	500 m
Number of sensor nodes	500
Initial Power	2 J
Maximum communication range	300 m
E_{elec}	50 nJ/bit
ϵ_{amp}	100 pJ/bit/m^2
Length of data packet	2000 bits

For multi-hop communication between clusters, we needed to determine the percentage of cluster heads. We defined the data collection rate as the portion of packets arriving at the sink node in all transmitted packets and investigated the change in the data collection rate by using LEACH and HEED as clustering methods with different cluster-head percentages. We found that the percentage of cluster heads has little influence on the data collection rate. In particular, we chose 20% as the cluster-head percentage because a high data collection rate can be maintained for a slightly longer time with this value.

4.3 Theoretic analysis for power consumption

Power consumption is a crucial metric in sensor networks whose power sources are severely constrained. Because the network's purpose is gathering data, all data generated by the sensor nodes are drawn into the sink node. Therefore, sensor nodes near the sink node consume a large amount of power in relaying data and power consumption imbalances from region to region are inevitable in large-scale sensor networks relying on multi-hop communication. Furthermore, this means we should evaluate the power consumption of smaller regions rather than that of the whole area, as other studies do. To do so, we analytically derived the energy consumption of the whole monitoring region and then show the region-specific power consumptions, by dividing up the whole monitoring area into regions based on the distance from the sink node.

We have already proposed a method for analytically deriving the power consumption of multi-hop communications between clusters [8]. Our approach extends the method of [13] to multi-hop communication between clusters. The analytical model is based on two assumptions. First, the transmission distances of sensor nodes and cluster-heads are constant (i.e., no transmission power control). Second, the relay in the routing between cluster heads is to the adjacent cluster in the direction of the sink node.

We can derive the power consumption when a node acts as an ordinary sensor node by first assuming that the clusters form a circle of radius l. Here, we consider the case in which the cluster head collects k bits generated in all sensor nodes within a cluster. The number of sensor nodes in each cluster, N_c, is

$$N_c = \frac{N}{N_{CH}} \tag{7}$$

where N is the number of sensor nodes in a whole monitoring region and N_{CH} is the number of cluster heads. Here we define r as the communication range of a sensor node. Thus, the number $n(h)$ of sensor nodes left for h hops to their cluster head can be calculated as follows:

$$n(h) = N_c \frac{r^2}{l^2} \left\{h^2 - (h-1)^2\right\} \tag{8}$$

Moreover, assuming that nodes with the same hops from the cluster head share the load for relaying data equally, the number $l(h)$ of transmissions for a node h hops from the cluster head can be calculated as follows:

$$l(h) = \frac{\sum_{i=h}^{\lceil \frac{l}{r} \rceil} n(i)}{n(h)} = \frac{N_c \left\{1 - (h-1)^2 \frac{r^2}{l^2}\right\}}{n(h)} \tag{9}$$

Therefore, we derive the total number of transmissions in a cluster, x_c, as

$$x_c = \sum_{h=1}^{\lceil \frac{l}{r} \rceil} N_c \left\{1 - (h-1)^2 \frac{r^2}{l^2}\right\} \tag{10}$$

Now we can obtain the total energy E_c required for the cluster head of a certain cluster to collect data as follows:

$$E_c = x_c(E_{elec}k + \epsilon_{amp}kr^2) + x_c E_{elec}k \tag{11}$$

Next, we derive the energy consumption when the node acts as a cluster head. To do this, we consider cluster heads in a annular domain of width r_{CH} and whose distance from a sink node is between $d - r_{CH}/2$ and $d + r_{CH}/2$ as shown in Fig. 3, where r_{CH} is the transmission range of the cluster head. Since most relaying to a cluster head is close to the sink node, we consider that the probability of relaying between cluster heads in this annular domain is very small. Moreover, data from a cluster head outside of the annular domain cannot jump over the domain and be transmitted to a cluster head inside it. Therefore, these cluster heads in the annular domain will receive the data generated by all the sensor nodes located outside that domain and will relay it to the sink node or a cluster head closer to the sink node.

Since the areal density of sensor nodes is $N/\pi R^2$ and the areal density of cluster heads is $N_{CH}/\pi R^2$, the amount of data from sensor nodes outside the range of $d + r_{CH}/2 < R$ is

$$\frac{R^2 - (d + r_{CH}/2)^2}{R^2} Nk \tag{12}$$

Moreover, although the number of cluster heads in a annular domain is

$$\frac{(d + r_{CH}/2)^2 - (d - r_{CH}/2)^2}{R^2} N_{CH} \tag{13}$$

for simplicity, the data relayed from outside that domain shall be equally divided among the cluster heads in the domain. Furthermore, each cluster-head combines the data generated by N/N_{CH} nodes in its cluster and transmits that data to the sink node or a cluster head closer to the sink node.

On the other hand, a cluster head in the range of $d + r_{CH}/2 > R$ does not relay data farther from the sink node. It receives only the data generated in the sensor nodes within its cluster and transmits it to the next cluster-head. Moreover, if a cluster is a circular area of radius l, in order that a cluster may stick out of the sensing area, the number of sensor nodes in such a cluster is calculated from the following area. The area of a circular cluster that overlaps in sensing area can be derived as follows:

$$l^2 \left\{\pi - \arccos\left(\frac{x-d}{l}\right)\right\} + R^2 \arccos\left(\frac{x}{R}\right) - dy \tag{14}$$

where

$$x = \frac{R^2 - l^2 + d^2}{2d}, y = \sqrt{R^2 - x^2}$$

and the number of sensor nodes can be approximately obtained. Now we can obtain the power consumptions of a sensor node and a cluster head. The results of the analysis and simulation experiments are shown in Fig. 4. These results were obtained for 2000 sensor nodes, a 300-m-radius monitoring region, and a 20-m communication range for the sensor node. Although the parameters are different from the ones in Table. 1, the simulation results for $r_{CH} = 100$ and those of the analysis are in good agreement when transmission power control is not performed and the communication range is fixed. On the other hand, the simulation values are slightly larger than the analytical ones when $r_{CH} = 40$. Although our analytical model assumes the clusters contains equal numbers of sensor nodes, in practice, the number of sensor nodes in a cluster will vary. Especially when r_{CH} is small, some clusters

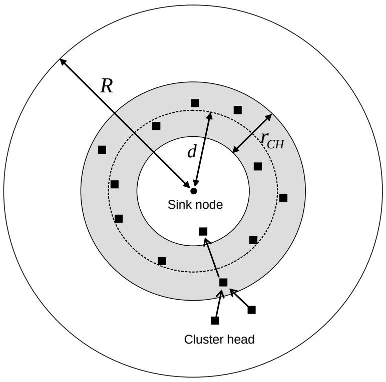


Fig. 3 Data generated outside of the annular area is relayed via the cluster heads in that area to the inner area.

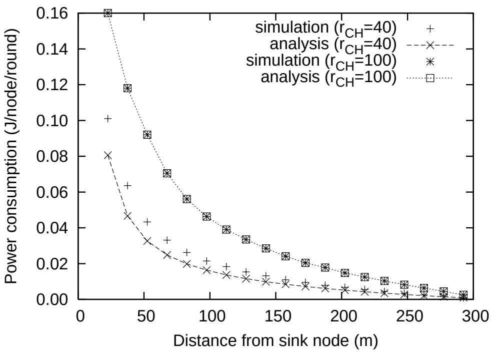


Fig. 4 Relationship between distance from sink node and power consumption of sensor nodes derived by analysis and simulation.

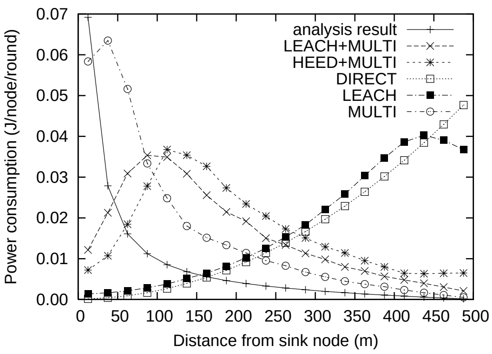


Fig. 5 Power consumptions of various routing protocols.

might consist of only a cluster head. Therefore, in comparison with the analysis, the simulation experiments used more cluster heads that consumes more power in relaying and broadcasting cluster advertisements. The above results indicate that our implementation of the simulation programs worked correctly.

4.4 Evaluation of power consumption

In large-scale sensor networks affected by power consumption imbalances, we must evaluate individual monitoring regions, not simply the whole region. To do so, we evaluated the per-node power consumption by using the above analysis and simulation experiments. Since transmission power control is disabled in our analytical model, we obtained the average transmission distances of LEACH+MULTI and used them in the analysis. This result is plotted in Fig. 5. If one chooses to simply repeat data relays, as mentioned in [8], the power consumption grows larger closer to the sink node because transmission power is not controlled. In the analytical model, transmissions could consume excessive power. On the other hand, the simulation experiments consumed just enough power to communicate with a next-hop node. Comparing results of LEACH+MULTI and the analysis, we see that their power consumptions near the sink node vastly differ despite the equality of their average communication distances. This result is attributed to the simulation model's suppression of transmit power consumption around the sink node. Thus, transmission power control not only suppresses unnecessary power consumption but also reduces power consumption imbalances.

In the simulation experiments, if a region exists where sensor nodes cannot transmit packets to the sink node directly, it would be difficult to compare routing protocols because not all packets can reach the sink node. Therefore, we set the maximum communication range to 500 m for DIRECT and LEACH.

MULTI generates a region 25 – 50 m from the sink node in which the sensor node consumes the largest amount of power (see Fig. 5). The sensor nodes in this region must relay enormous numbers of packets from other sensor nodes farther away from the sink node. LEACH+MULTI and HEED+MULTI reduce power consumption since they shift the role of cluster heads between sensor nodes so that the relay load is dispersed among them. The region where the largest amount of power is consumed is 100 m from the sink node. The main reason for this is the probability of cluster heads being located inside this area is low, so cluster heads in this area must communicate with the sink node directly without relays.

We also conducted simulations to investigate the power consumption characteristics in a monitoring region with a 60 m radius to determine whether multi-hop communication between clusters is applicable to small networks. In small networks, less power is consumed when the percentage of cluster heads is low. We set the percentage to 5% of the total number of sensor nodes. Moreover, considering the size of the monitoring region, we limited the range of cluster-head advertisements to within a circle with a 60 m radius. The results are shown in Fig. 6. The figure indicates that LEACH and DIRECT consume minimal power because of the small monitoring region. In such a region, the power consumption of the transmit amplifier is vanishingly

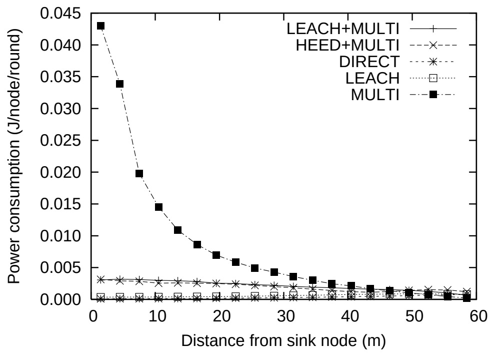


Fig. 6 Power consumption characteristics in small monitoring region

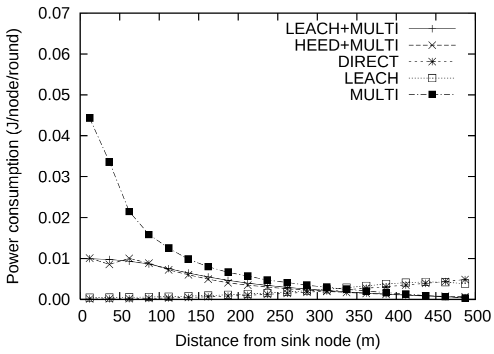


Fig. 7 Power consumption with $\epsilon_{amp} = 10$ [pJ/bit/m^2]

Table 2 Average hop counts

HEED+MULTI	LEACH+MULTI	MULTI
7.061	7.450	15.790

small, and this means a relatively increase in power for running radio circuitry. In LEACH and DIRECT, the number of transmissions for a non-cluster-head node (called regular node) is only 1, as contrasted with the situation that many transmissions may be needed in multi-hop communication. While HEED+MULTI and LEACH+MULTI consume less power than MULTI does, they need more than triple the power of LEACH or DIRECT. As a result, multi-hop communication between clusters does not show a power consumption advantage in small networks.

We used a first-order radio model as the power consumption model and the same parameter values described in [5]. On the one hand, this model and the parameter values are in wide use, On the other hand, there have been some experimental measurements of the sensor nodes' power consumption. For example of the latter, [14] describes the power consumption of the radio in Wireless Integrated Networks Systems (WINS). This research does not fit a first-order radio model, and its results cannot be applied to our research without modification. However, what affect the performance of the network is the ratio E_{elec} to ϵ_{amp}. Therefore, we restricted our consideration to the result that the proportion of transmission power to reception power is 2:1 [14] and adjusted ϵ_{amp} to this proportion while keeping E_{elec} constant. The average communication distance of LEACH+MULTI is about 70 m. In this case, ϵ_{amp} should be set to 10 [pJ/bit/m^2]. Figure 7 shows the resulting distribution of power consumption with these parameters. The figure is similar to Fig. 6, and LEACH and DIRECT have minimal power consumption. However, given a vast monitoring region, multi-hop communication is necessary. The power consumption of multi-hop communication is considerably suppressed by clustering. That is, reducing ϵ_{amp} leads to relative power consumption increase for running radio circuitry, as in the case of the small monitoring region, and suppressing the number of transmissions has higher priority than keeping the transmission distance short. Because LEACH and DIRECT employ 1-hop communication to the sink node, the power consumed by circuitry is extremely small and power for long-distance communication has a small impact because of the small ϵ_{amp}. Table 2 lists the average hop counts from a sensor node to the sink node. The hop counts of LEACH+MULTI and HEED+MULTI are lower than MULTI, and this decreases the overall power consumption.

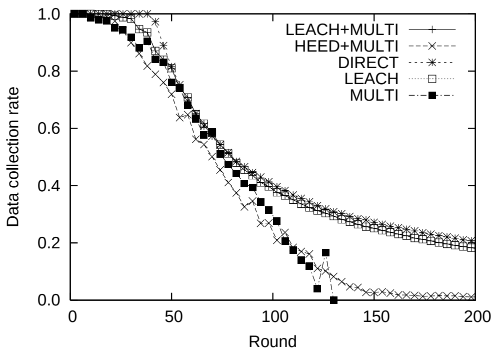


Fig. 8 Deterioration of data collection rate due to power depletion.

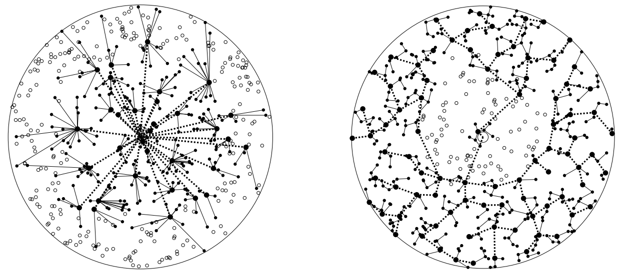
4.5 Deterioration of data collection rate due to power depletion

Maintaining a high data collection rate is important because the data collection rate determines the lifetime of sensor networks. The data collection rate per round is shown in Fig. 8. In this figure, each curve, which represents the data collection rate of a different routing method, behaves similarly until about 80 rounds. However, the positions of exhausted sensor nodes make a difference, as shown in Fig. 5. To make this difference clearly understandable, we visualized these distributions of exhausted sensor nodes of LEACH and HEED+MULTI in Figs. 9(a) and 9(b). These figures represent the situation of a sensor network at 80 rounds. The filled circle at the center of the monitoring region is the sink node. Small filled circles and open circles represent sensor nodes that have residual power or exhausted sensor nodes, respectively. Solid lines are the communication links within a cluster, and dotted lines are the communication links from a cluster head to the sink node and between cluster heads. In LEACH and DIRECT, the sensor nodes on the fringe of the monitoring region exhaust their power first, since these nodes must transmit packets to the distant sink node in a single hop. Due to their power depletion, the packet collection rate starts to decrease early on. As time goes by, however, the packet collection rate curves for LEACH and DIRECT decrease slower than the other methods'. This indicates a situation where only sensor nodes near the sink node have residual power, as in Fig. 9(a). These sensor nodes can operate for many rounds since the transmission power needed to reach the sink node is comparatively small. However, a non-uniform distribution of the residual power is undesirable in sensor networks because of the need to collect data over the entire monitoring region. As shown in Fig. 9(b), HEED+MULTI also has an unbalanced residual power distribution. In this routing method, the sensor nodes near the sink node deplete their power early. The same can be said for the other two routing methods using multi-hop communications (i.e., LEACH+MULTI and MULTI). As power becomes depleted near the sink node, it becomes more and more difficult to gather information. In addition, obtaining information about the fringe of the monitoring region also becomes difficult. Because of these problems, the data collection rate will fall sharply.

It seems like that LEACH and DIRECT are preferable to other methods using multi-hop communication. However, the area where LEACH and DIRECT can maintain higher rates is not of practical use. As for multi-hop communication, we adopted very a simple routing using only distance as a criterion in this thesis but if we adopt more intelligent routing that allows for other criteria such as residual power, there is still room for increasing the data collection rate. The routing problem is one of the hottest topics in the fields of sensor networks and ad-hoc networks.

4.6 Data collection time

The data collection time is important consideration for applications requiring swift data collection in large-scale sensor networks. Fig. 10 is a graph of the data collection time. The x-axis shows the timeslots, and the y-axis indicates the cumulative number of packets which arrived at the sink node. The results are similar to those of the monitoring region with the 60 m radius. Because we set the number of sensor nodes to 500, the number of packets the sink node receives is 500. Therefore, the minimum time to collect all the data is 500 timeslots. LEACH and DIRECT take the minimum time and are optimal. In DIRECT, 1 timeslot is simply assigned to each sensor node to avoid transmitting data simultaneously with other sensor nodes. Regarding LEACH, while a cluster head communicates with the sink node, the other cluster heads can collect data from sensor nodes in their clusters. Consequently, a transmission schedule that is free of interference and has an optimal data collection time can be constructed for it.



(a) LEACH (b) HEED+MULTI

Fig. 9 Situations at 80 rounds in (a) LEACH and (b) HEED+MULTI.

In sensor networks using multi-hop communication, the amount of data arriving at the sink node does not grow linearly with time, regardless of the existence of clusters. The reason for this is that not only do packets generated by sensor nodes far away from the sink node take a long time to reach the sink node but also a waiting period is needed to avoid interference. The data collection time with clusters decreases in comparison with the case without clusters. The key reason for this is that we use different two channels for intra-cluster communication and inter-cluster communication, respectively. On the one hand, if we use one channel for the entire network, many sensor nodes around the cluster heads have to delay their transmissions frequently to avoid interfering with inter-cluster communications because the communication between cluster heads has a large interference range. On the other hand, if we use two channels as described above, no interference occurs between intra-cluster and inter-cluster communications. The two transmissions can be performed in parallel. This two-tiered form of multi-hop communication between clusters gathers data quicker than one-tiered multi-hop does.

4.7 Summary of the characteristics

In this section, we investigate the characteristics of multi-hop communication in clustered sensor networks from a broad perspective. The investigation shows that this form of communication is effective for large-scale

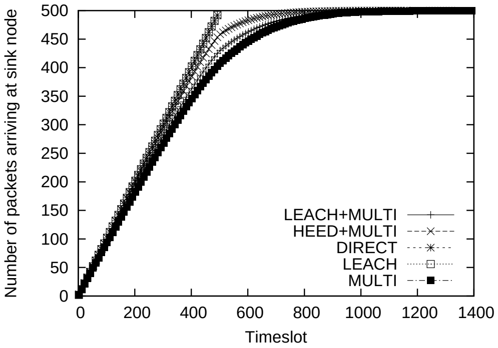


Fig. 10 Data collection time for different routing protocols.

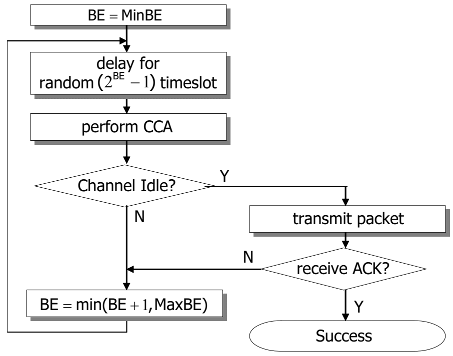


Fig. 11 Transmission algorithm of CSMA/CA.

sensor networks. First, we can say that 1-hop transmission to the sink node is difficult in a vast monitoring region because the transmission distance of the radio is limited. Therefore, multi-hop communication is absolutely essential. In terms of the highest per-region power consumption, multi-hop communication between clusters cuts power consumption in comparison to the cases without clustering by 56% (LEACH+MULTI) and 58% (HEED+MULTI) when $\epsilon_{amp} = 100$ [pJ/bit/m^2], by 22% (LEACH+MULTI) and 23% (HEED+MULTI) when $\epsilon_{amp} = 10$ [pJ/bit/m^2]. This has a large beneficial effect on sensor networks whose power sources are severely constrained. Moreover, LEACH+MULTI reduces the data collection time by 10% and HEED+MULTI reduces it by about 20% compared with MULTI. These values indicate that the channel configuration used in this research will reduce the data collection time as well as the power consumption.

On the other hand, in small sensor networks, multi-hop communication between clusters has no advantage regarding power consumption even with respect to DIRECT, which is the simplest communication form. Therefore, we must seriously consider its domain of applicability. For example, LEACH and DIRECT achieve better performance in terms of both power consumption and data collection time. Multi-hop communication between clusters shows its advantages in large-scale networks. Thus, in a large monitoring region where direct transmission is impossible, we recommend using multi-hop communication between clusters.

5. Applying CSMA/CA to sensor networks

In the previous section, we evaluated the fundamental characteristics of multi-hop communication in clustered multi-hop sensor networks using TDMA based on an ideal transmission schedule and showed the benefits of such communication in large-scale sensor networks. In a practical sense, however, constructing such a transmission schedule is an improbable scenario because it needs positional information on all sensor nodes. To use this TDMA, an enormous amount of information has to be exchanged and this consumes great deal of time. In a realistic situation, frequent interference is unavoidable and an ideal schedule cannot be used. We think that it would be useful for laying sensor networks to know how much interference deteriorates performance. Thus, we compared TDMA communication using the ideal schedule (as described in the previous section) with CSMA/CA communication, which can deal with interference. Applying CSMA/CA, we quantitatively clarify the increases of power consumption and data collection time caused by retransmissions, compared with the ideal TDMA.

The CSMA/CA mechanism we used is based on IEEE 802.15.4 [16]. Each sensor node maintains a variable backoff exponent (*BE*). The *BE*s are used for determining the length of random waiting times before evaluating the status of a channel. The transmission algorithm is shown in Fig. 11. In CSMA/CA, the maximum number of backoffs is provided as a threshold, and if the number of backoffs reaches this threshold, the sensor node aborts its transmission attempt. Here, however, we assume that the sensor node backs off as many times as needed to account for the influence of retransmissions caused by data interference. The values of *MinBE* and *MaxBE* are set to 3 and 5 respectively. These values are the defaults of IEEE 802.15.4.

Retransmissions are needed even when using a CSMA/CA mechanism because of the hidden terminal problem. When this problem occurs, a sensor node receives multiple packets in a timeslot (of course, these packets cannot be read correctly) and the node does not notify the sending nodes that it received them. In this process, not only the sending nodes but also the receiving node consumes power. We take into account this in the power consumption evaluation. In addition, back off must be done before carrier sensing in CSMA/CA. Consequently, the power consumption and data collection time will increase. We investigate this situation when it occurs in HEED+MULTI with CSMA/CA and HEED+MULTI with TDMA based on an ideal transmission schedule.

5.1 Power consumption

Unlike TDMA communication using an ideal schedule, CSMA/CA communication is susceptible to interference due to hidden terminal problems. Fig. 12 compares the power consumption levels of a sensor node. The power consumption of sensor networks with CSMA/CA is larger than that of sensor networks with TDMA using an ideal transmission schedule. This is because of retransmission caused by interference. CSMA/CA did not completely avoid data interference, as noted above, and in this case, the sensor node transmitted data at least twice. The first transmission caused interference. The closer the sensor nodes are to the sink node, the more data is relayed. The probability of interference affecting the transmissions of these nodes thus increases. The sensor node one hop from the sink node (the last-hop node) relays the largest number of packets, and the smaller the percentage of the cluster-heads is, the longer the distance between last-hop nodes and the sink node becomes. Therefore, the transmission power of the last-hop node increased. These factors cause the power consumption in the region 75 – 100 m from the sink node to be maximal when the percentage of the cluster heads is 20%. This region is where the most power is needed by TDMA using the positional information on all sensor nodes. The main problem is that it becomes impossible to collect data from the sensor nodes in as their power is depleted. Therefore, the power consumption distribution should be obtained.

We define E_{TDMA} to be the power consumption of TDMA using a transmission schedule that avoids interference, and $E_{CSMA/CA}$ as the power consumption with CSMA/CA. We also define the percentage increases as $\frac{E_{CSMA/CA}-E_{TDMA}}{E_{TDMA}}$. By calculating the average percentage increases over all monitoring regions, the power consumption of CSMA/CA increases by 12% compared with that of TDMA. These results indicate that the power consumption of CSMA/CA is not much worse than that of interference-free TDMA.

5.2 Data collection time

Power consumption is not only the performance aspect that is degraded by interference. Data collection time is also worsened by backoff in CSMA/CA. We compared TDMA communication with CSMA/CA communication as to their deterioration of data collection time.

A backoff period is necessary before status of a channel in CSMA/CA can be evaluated, and the data collection time may become longer than TDMA. Fig. 13 compares the data collection time. The period to collect 90% of the packets is 3.7 times longer in CSMA/CA. In large-scale sensor networks where many sensor nodes are densely deployed, when a sensor node tries to transmit a packet, channels are often busy with the transmissions of other sensor nodes. Therefore, back off has to be done many times. Moreover, the more transmissions there are, the longer the backoff period will be, because of the truncated binary exponential backoff of IEEE 802.15.4, and many transmissions occurs near the sink node. All the data must go through the area, and this is one reason that data collection time increases.

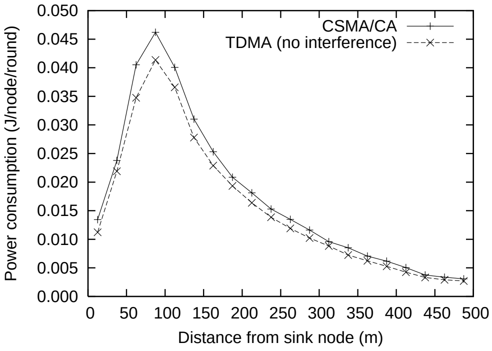


Fig. 12 Power consumptions per node of CSMA/CA and TDMA with interference-free scheduling.

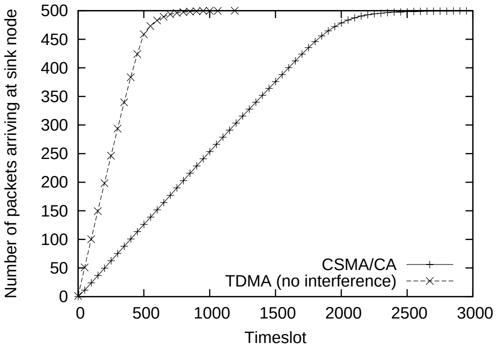


Fig. 13 Comparison of data collection rates.

5.3 Discussion

The time for CSMA/CA to gather the same amount of data is 3.7 times that of the ideal TDMA without interference. This seems like CSMA/CA will face frequent interference with multi-hop communication between clusters. Yet CSMA/CA's power consumption is only 12% larger than the ideal TDMA without interference. To determine the reason for it, we counted the number of backoffs in inter-cluster communication and in intra-cluster communication (see Table 3). It is evident from the table that most backoffs are performed to avoid interference and backoffs due to actual interference are just a 5% of the total backoffs. This accounts for suppression of increase in power consumption. On the other hand, the data collection rate with CSMA/CA is significantly deteriorated compared

Table 3 Relationship between the number of backoff and that of interference which occurs retransmission on each channels

	(1) Backoff	(2) Interference	(2)/(1)×100
Inter-cluster	2640	133	5.02
Intra-cluster	6459	269	4.17

with that of TDMA. Although a power source is a crucial resource for sensor networks, this trade-off between power consumption and data collection rate indicates that sensor networks with a large hop counts will have to be carefully planned if they are to have real-time performance.

6. Conclusion

Multi-hop communication between clusters is preferable for data collection in large-scale sensor networks. We focused the fundamental characteristics of multi-hop communication in a large-scale clustered sensor network and clarified three characteristics (power consumption distribution, the effect of the distribution on data collection rate, and data collection time) with a TDMA scheme based on an ideal interference-free transmission schedule. Compared with simple multi-hop communication, multi-hop communication between clusters can alleviate the concentration of the relay load and reduce the data collection time. We used a relatively simple routing method in this research, so we believe that the use of more intelligent routing methods will further improve the performance of multi-hop communication between clusters.

Understanding how much performance degradation arises from interference is another important issue. We compared CSMA/CA, which tolerates interference, with TDMA and showed that CSMA/CA takes 3.7 times longer to gather equal amounts of data and that it suppresses power consumption by only 12% in comparison with ideal TDMA. These results can be attributed to the backoff time of CSMA/CA and retransmissions caused by interference.

This research identified some ways of improving the performance of multi-hop communication between clusters. These will be the starting points for future work. As we have indicated, the relay load concentrates on the sensor nodes around the sink node. If the power of these nodes is depleted, the packets going through them do not arrive at the sink node and the data collection rate deteriorates rapidly. Therefore, we think that source routing is not as effective as next-hop routing in sensor networks. In addition, it is essential to smooth out power consumption imbalances. In the future, we will devise routing algorithms for multi-hop communication between clusters that include the above considerations.

Acknowledgements

The authors would like to thank the anonymous reviewers and Dr. Kenji Leibnitz for his helpful comments and suggestions. This research was partly supported through a grant titled "New Information Technologies for Building a Networked Symbiosis Environment" (The 21st Century Center of Excellence Program) and a Grant-in-Aid for Scientific Research (C) 17500043 of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

[1] H. Chan and A. Perrig, "ACE: An emergent algorithm for highly uniform cluster formation," Proc. the First European Workshop on Wireless Sensor Networks, Jan. 2004.
[2] W. Choi, P. Shah, and S.K. Das, "A framework for energy-saving data gathering using two-phase clustering in wireless sensor networks," Proc. the First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, pp.203–212, Aug. 2004.
[3] B.J. Culpepper, L. Dung, and M. Moh, "Design and analysis of hybrid indirect transmissions (HIT) for data gathering in wireless micro sensor networks," ACM Mobile Computing and Communications Review, vol.8, no.1, pp.61–83, Jan. 2004.
[4] S. Ghiasi, A. Srivastava, X. Yang, and M. Sarrafzadeh, "Optimal energy aware clustering in sensor networks," Sensors Magazine, vol.2, pp.258–269, July 2002.
[5] W.R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," Proc. the Hawaii International Conference on System Sciences, pp.1–10, Jan. 2000.
[6] S. Lindsey, C. Raghavendra, and K.M. Sivalingam, "Data gathering algorithms in sensor networks using energy metrics," IEEE Transactions on Parallel and Distributed Systems, vol.13, no.9, pp.924–935, Sept. 2002.
[7] O. Younis and S. Fahmy, "Distributed clustering in ad-hoc sensor networks: A, hybrid, energy-efficient approach," Proc. IEEE INFOCOM 2004, pp.629–640, March 2004.
[8] M. Sugano, Y. Kiri, and M. Murata, "Performance analysis of large-scale wireless sensor network architecture with multi-cluster configuration," Proc. the IASTED International Conference on Networks and Communication Systems, pp.188–193, March 2006.
[9] A.A. Papadopoulos and J.A. McCann, "Connectionless probabilistic (CoP) routing: an efficient protocol for mobile wireless ad-hoc sensor networks," Proc. the 24th International Performance Computing and Communications Conference, pp.73–77, April 2005.
[10] J. Neander, E. Hansen, M. Nolin, and M. Björkman, "Asymmetric multihop communication in large sensor networks," Proc. International Symposium on Wireless Pervasive Computing 2006, Jan. 2006.
[11] S. Soro and W.B. Heinzelman, "Prolonging the lifetime of wireless sensor networks via unequal clustering," Proc. the 19th IEEE International Parallel and Distributed Processing Symposium, April 2005.
[12] T. Shu, M. Krunz, and S. Vrudhula, "Power balanced coverage-time optimization for clustered wireless sensor networks," Proc. ACM MobiHoc '05, pp.111–120, May 2005.
[13] E.J. Duarte-Melo and M. Liu, "Energy efficiency of many-to-one communications in wireless networks," Proc. IEEE

45th Midwest Symposium on Circuits and Systems, Aug. 2002.
[14] A. Savvides, S. Park, and M.B. Srivastava, "On modeling networks of wireless microsensors," Proc. Joint international conference on Measurement and modeling of computer systems, pp.318–319, June 2001.
[15] L.M. Feeney, B. Ahlgren, and A. Westerlund, "Investigating the energy consumption of a wireless network interface in an ad hoc networking environment," Proc. IEEE INFOCOM 2001, pp.1548–1557, April 2001.
[16] IEEE, 802.15.4–2003 IEEE Standard for Information Technology–Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low–Rate Wireless Personal Area Networks (LR-WPANs), Oct. 2003.

Yuichi Kiri received his B.E. from Osaka University, Japan in 2006. Since April 2006, he has been a graduate student at the Graduate School of Information Science and Technology, Osaka University. His research interests include the design, implementation, and simulation of wireless sensor networks and mobile wireless networks.

Masashi Sugano received the B.E., M.E., and D.E. degrees in Information and Computer Sciences from Osaka University, Japan, in 1986, 1988, and 1993, respectively. In 1988, he joined Mita Industrial Co., Ltd. (currently, Kyocera Mita Corporation) as a Researcher. From 1996 to 2003, he was an Associate Professor in Osaka Prefecture College of Health Sciences. From 2003 to 2005, he was with the Faculty of Comprehensive Rehabilitation, Osaka Prefecture College of Nursing, and since April 2005, he has been an Associate Professor of the School of Comprehensive Rehabilitation, Osaka Prefecture University. His research interests include performance evaluations of computer communication networks, network reliability, and ad hoc and sensor network systems. He is a member of IEEE, ACM, IEICE and IPSJ.

Masayuki Murata received the M.E. and D.E. degrees in Information and Computer Sciences from Osaka University, Japan, in 1984 and 1988, respectively. In April 1984, he joined IBM Japan's Tokyo Research Laboratory, as a Researcher. From September 1987 to January 1989, he was an Assistant Professor with the Computation Center, Osaka University. In February 1989, he moved to the Department of Information and Computer Sciences, Faculty of Engineering Science, Osaka University. In April 1999, he became a Professor of Osaka University, and since April 2004, he has been with the Graduate School of Information Science and Technology, Osaka University. He has contributed more than four hundred and fifty papers to international and domestic journals and conferences. His research interests include computer communication networks and performance modeling and evaluation. He is a member of IEEE, ACM, The Internet Society, and IPSJ.