Probabilistic Data Collection Protocols for Energy Harvesting Wireless Sensor Networks

Masaya Yoshida∗, Tomoya Kitani∗, Masaki Bandai‡, Takashi Watanabe∗, Pai Chou¶, and Winston Khoon Guan Seah^
∗Shizuoka University, Japan; ‡Sophia University, Japan;
¶University of California, Irvine, USA; ^Victoria University of Wellington, New Zealand

***Abstract*—Energy harvesting has been studied as a candidate for powering next generation wireless sensor networks. Ambient energy sources that can be harvested to power sensors include solar, vibration, heat and wind. However, sensor nodes powered by energy harvesting devices cannot always communicate with other nodes because the energy harvesting devices of the same footprint as sensor are unable to provide a stable supply of energy. As a result, a node cannot know whether its neighboring nodes have enough energy to receive a data packet that it has transmitted. During the process of relaying the packet, each additional hop increases the overall probability of losing the packet. In this paper, we propose two data collection protocols for energy harvesting wireless sensor networks called the Probabilistic ReTransmission protocol (PRT) and PRT with Collision Consideration (PRT-CC). The idea is to derive the number of times to retransmit a packet based on the reception probability and the active intervals computed by the receivers themselves while, in PRT-CC, each node computes the reception probability with packet collision consideration. In addition, the use of acknowledgements (ACKs) provides another means to detect packet loss. The performance evaluation shows that the proposed protocols are able to achieve higher delivery ratio than the previous work, namely, Geographic Routing with Duplicate Detection (GR-DD) and GR-DD with Retransmission (GR-DD-RT). As nodes within a locality tend to exhibit similarity in their energy harvesting profile, we also validated our protocols under such scenarios.**

***Index Terms*—wireless sensor network, energy harvesting, data collection, ambient**

I. Introduction

Wireless sensor networks (WSNs) are networks consisting of compact devices called sensor nodes that collect data such as temperature and humidity over the target area. Sensing data are relayed via wireless links between nodes using a multi-hop communication protocol. Most of studies on WSNs have been motivated by their wide range of applications, ranging from natural environmental monitoring and disaster data collection to security systems.

Since WSN nodes may be deployed in large numbers and mostly run on batteries, the maintenance of nodes is required, including replacing batteries that have been depleted of energy. However, largely-scale maintenance costs are high or even prohibitive. For example, it is very difficult, even if possible, to reach nodes frequently in remote locations such as dense natural vegetation or mountainous regions. Thus, to extend the lifetime of WSNs, one of the important issues is saving power while providing the desired quality of communications. To address this issue, various protocols for power-efficient data collection have been proposed [1][2].

Recently, energy harvesting has received growing attention

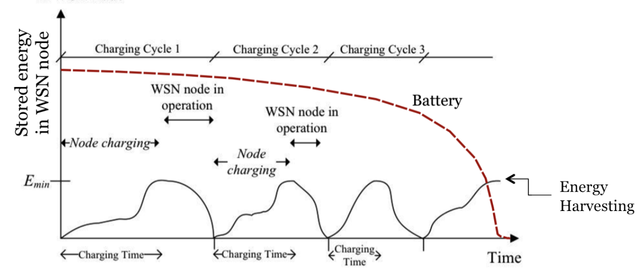


Fig. 1. Temporal change of energy level of a battery and an energy harvesting source

in next-generation WSNs [3] leading to the design and development of energy harvesting WSNs [4][5]. Energy harvesting entails converting forms of ambient energy from the environment, such as light, thermal differential and vibration into energy. Such renewable sources are expected to reduce the need for frequent maintenance, thus enabling sustainable operation of WSNs. However, due to the unpredictable supply of harvested energy, it is difficult to determine the operating state of neighboring nodes. For this reason, conventional protocols that assume a reliable (albeit limited) energy source provided by batteries is not applicable to WSNs utilizing energy harvesting for power. Therefore, protocols need to be designed with energy harvesting assumptions in mind.

This paper proposes two new protocols, called Probabilistic ReTransmission (PRT) and PRT with packet Collision Consideration (PRT-CC), with the goal of achieving high efficiency and reliability in data collection in the presence of unsteady power supplied by energy harvesting. In addition, we consider the use of acknowledgements (ACKs) to provide another means to detect packet loss. We evaluate the effectiveness of our protocols by simulation and results confirm the efficacy of the proposed approach.

II. Related Work

WSNs can potentially contain large numbers of nodes. The size of nodes are usually small for low cost and ease of deployment. Thus, energy harvesting devices, which must also be small, are usually unable to sustain the continuous operation of most sensor nodes. Figure 1 shows an example of the energy level of a battery and an energy harvesting device of similar size.

A network with such unsteady power supply conditions cannot operate at all times. Furthermore, charging takes a variable amount of time depending on the environment, making it difficult to predict the charging state of the surrounding nodes. As a result, it is difficult to determine the best path to the sink node for data transmission. Such characteristics specific to WSNs powered by ambient energy harvesting are very different from those traditionally powered by batteries, making it difficult to adopt previously designed WSN protocols. Thus, it is necessary to develop new protocols that considers media access control, routing, topology maintenance, sleep control, and power management, including charging control, in order for WSNs to operate efficiently when powered by harvested energy sources.

One of related work is an energy-harvesting WSN for a railway monitoring system [6]. In [7], the authors introduce mobility to the power control. In [8], abstractions are developed to characterize the complex time varying nature of the harvesting sources, and it has been shown by [6], [7] and [8] that power control schemes can enable limited power to be used effectively. In [9], the authors study the rate assignment problem for rechargeable sensor networks and propose protocols to compute the optimal rate assignment. Two power control metrics have been developed for nodes powered by energy harvesting systems [10] which depend on the average queue length (the number of unsent messages) and average data loss rate (i.e., the amount of data that a node cannot receive during its sleeping/charging interval). In [11], a bridge monitoring application is studied with specific emphasis on the placement of the nodes and a data collection protocol optimized for the network topology and energy harvesting efficiency, while MAC protocols that can be used in energy-harvesting WSNs are studied and analyzed in [12]. In [13], the energy harvesting process is modeled as a discrete process with "packets of energy" arriving at specific time instances, and an optimal solution is provided for short-term throughput maximization. The solution is validated analytically to show that it is advantageous to use optimum power allocation for transmitters with limited battery capacity or highly variable energy harvesting rates. The problem of maximizing the global sustained sampling rate is addressed in [14] and the solar-aware distributed flow (SDF) approach has been proposed; SDF aims to maximize the sampling rate while maintaining energy neutral operation.

As we will describe the problem in detail in section III, the instability of power supply arising from energy harvesting incurs packet loss resulting in low packet delivery rate. To overcome this problem, Geographic Routing with Duplicate Detection (GR-DD) and Geographic Routing with Duplicate Detection and Retransmission (GR-DD-RT) have been proposed [11]. In the GR-DD protocol, if a node receives the same data packet multiple times, then the latter packets are discarded to reduce unnecessary power consumption. GR-DD-RT is an extension of GR-DD, where if a node is fully charged and is ready to transmit but there is no unsent packet in the queue, it retransmits the previously sent packet (i.e., nodes repeatedly transmit as much as possible). These are simple protocols designed for use only in simple topologies like a linear topology, and therefore they cannot achieve or provide higher performance in some scenarios. In this paper, we consider an arbitrary network topology and develop the data collection protocols that can achieve the higher delivery ratio than these protocols.

III. System Model

Our model is based on the repeating charging-and-transmitting model as proposed in [11]. Each sensor node is powered by an energy harvesting source, and it performs sensing and data transmission. The time history of the stored energy on a node is shown in Fig. 2, and each node operates according to the finite state machine shown in Fig. 3. During deployment, all nodes are pre-charged and therefore are able to execute a simple neighbor discovery process that involves broadcasting their locations to one another. Additional nodes that are deployed later will also broadcast their locations to their neighbors using the same neighbor discovery process. Therefore, we assume that all nodes know their own location and where the sink is.

Each node operates in the three states: *charge*, *receive*, and *transmit*. After the receive or transmit state, a node returns to the charge state. At the charge state, a node charges up to the minimum amount of energy, denoted by E_f, that should be sufficient for it to receive and transmit a packet. It then enters the receive state for a constant period of time. A node senses or receives data during the receive period. The amount of energy consumed by sensing is typically much lower than that for wireless communications. Therefore, we have accounted for the (small) amount consumed by sensing in the energy used by the node in the receive state. The receiving window, denoted by t_{rx}, should include the whole period of the packet transmission time, denoted by t_{tx}. Thus, t_{rx} must be greater than t_{tx}, and here we set t_{rx} to twice of t_{tx}, i.e. $t_{rx} = 2t_{tx}$ in common with [11]. The transmission time t_{tx} can be calculated by dividing the packet size s by the transmission rate α, namely

$$t_{tx} = \frac{s}{\alpha}.$$

Let P_{rx} and P_{tx} denote the receiving power consumption of a node and the transmitting power consumption of a node, respectively. Here E_f is set to the following equation as the energy required to perform one transmit and one receive as shown in Fig. 2,

$$E_f = P_{rx} \cdot t_{rx} + P_{tx} \cdot t_{tx}.$$

A received packet is put in the receiving node's queue. If there is a packet in the queue when the receive state ends and the channel is idle, then the node enters the transmit state and broadcasts the data packet at the head of the queue. At this time, any neighboring node that is currently in the receive state can receive the packet. After receiving and transmitting a packet, because the node has consumed its stored energy, it moves to the charge state. On the other hand, after the time the receiving state ends, if the queue is empty or the channel is not idle, then the node returns to the charge state until it charges up to E_f again. Each node repeats such cycles until the packet reaches the sink.

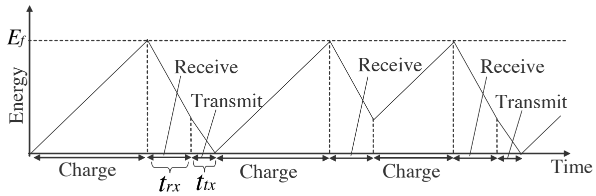


Fig. 2. The time history of the stored energy on the node

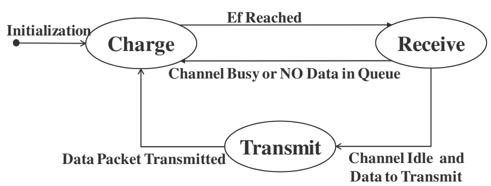


Fig. 3. The node operates according to the finite state machine

Even with the current state-of-the-art energy harvesting technology, if an energy harvesting device has the same (or even slightly larger) footprint as the wireless sensor node itself, the charging time can be significantly longer than the combined (receive and transmit) communication times. For instance, assuming that we use MICAz nodes with the recharge rate of 10mW and the packet size of 800 bits, the receiving time t_{rx} will be 6.4ms and the transmitting time t_{tx} will be 3.2ms, whereas the charging time is about 77ms. This means at any moment, most neighboring nodes are expected to be in the charge state.

We use this model to evaluate the simplest case where nodes can operate with small energy. It is possible to use the more complex model such as a node transmits and receives some packets in a cycle or a node charges and transmits/receives simultaneously. However, such a model makes the design of the sensor node more complex.

In this paper, we do not consider about processing overhead at each node because the duration for radio communication is much longer than that of processing and the energy consumption of the former is generally larger than that of the latter.

A. Data Collection Example

Let us consider an example of data collection (see Fig. 4) where each node has a different charge level. Node n_3 generates a data packet to be transmitted. Within the communication range of n_3, there are nodes n_1, n_2, n_4, n_5 and n_6. When a node is physically further away from the sink node than a sender, it does not enqueue its received packet but discards it instead, because it is unlikely for the packet to be sent closer to the sink. Therefore, in this case, nodes $\{n_1, n_2\}$ do not relay the packet from n_3. After node n_3 generated data, and recharged, if the channel is idle at that time, then n_3 broadcasts the packet of the data. Suppose that node n_4 is being recharged and nodes $\{n_5, n_6\}$ are in the receive state, then nodes $\{n_5, n_6\}$ can receive the packet, as shown in Fig. 4. In the same manner, nodes $\{n_5, n_6\}$ in turn broadcast the packet. The sink is assumed to

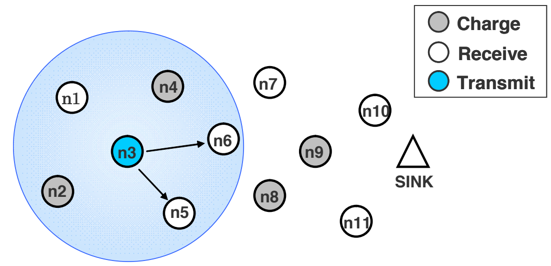


Fig. 4. Data collection example, where node n_3 transmits a packet

have steady power and is always in the receive state. In the manner described above, nodes relay the packets to the sink.

In another scenario, node n_3 generates another data packet and transmits it by broadcasting in the same manner as above. However, this time, suppose that all nodes $\{n_4, n_5, n_6\}$ are in charge state, and no node can receive the packet; consequently, the packet will be lost. This model assumes one-way communication where nodes transmit only data packets without transmitting ACK, or handshaking packets like Request-To-Send (RTS) and Clear-To-Send (CTS). Therefore, the sender node n_3 cannot know whether any of its neighboring nodes is able to receive the data packets it transmitted. As the packets are lost during relaying, the overall data collection rate deteriorates.

As the example shows, since the neighboring nodes are not always able to receive a packet during relaying, each additional hop increases the overall probability of losing the packet. Similarly, WSNs that take traditional approaches to the sleep control will suffer from the same problem, although a possible solution is to synchronize their sleep schedule. However, the synchronized approach cannot apply to energy harvesting WSNs due to the variable and unpredictable charging times of each node.

IV. Technical Approach

Our communication model is broadcast-based and a sender does not check whether any of its neighboring nodes actually receives the packet that it sent as shown in Section III. By retransmitting the same packet, a sender's neighbors have a higher probability of receiving the packet, but it needs additional energy for the retransmission and longer time to recharge. In order to decide how many times a sender retransmits a packet, we propose Probabilistic ReTransmission (PRT), by having each node calculate the probability of receiving a packet based on its own operating time and use this probability to decide whether to retransmit a packet or not. In addition, we propose PRT with Collision Consideration (PRT-CC), an extension of PRT which also considers the collision probability from hidden terminals of the sender.

The use of acknowledgements (ACKs) is common in many conventional networking protocols to explicitly notify the sender that a packet has been successfully received by the receiver. If the sender does not receive an ACK from the receiver within a given period after transmitting a packet, it will assume that the packet has been lost and retransmits it.

A. Probabilistic ReTransmission (PRT)

The idea of PRT is to derive the number of times to retransmit a packet based on the reception probability and the operating time computed by the receivers themselves. Let N_i denote the set of nodes within the communication range of node n_i, and let S_i denote the subset of N_i whose nodes are closer to the sink than n_i. For example, in the case of Fig. 4, $N_3 = \{n_1, n_2, n_4, n_5, n_6\}$, and $S_3 = \{n_4, n_5, n_6\}$. Each node can measure its own operating time with a timer. We assume that a node n_i can keep measuring a total charging time T_{chi}, a total receiving time T_{rxi}, and a total transmitting time T_{txi} during the period $[t-\tau, t]$, where t denotes a current time and $\tau(=T_{chi}+T_{rxi}+T_{txi})$ denotes a measurement period. If there is a node $n_h \in N_i \setminus S_i$ that broadcasts a packet to neighbor n_i in its communication range, the reception probability p_i can be approximated as follows:

$$p_i = \frac{T_{rxi}}{T_{chi}+T_{rxi}+T_{txi}} \cdot \frac{t_{rx}-t_{tx}}{t_{rx}}. \quad (1)$$

In Equation (1), T_{chi}, T_{txi} and T_{rxi} are measublack values whereas t_{tx} and t_{rx} defined in Section III are given values. The first half of the right side means the ratio of being in the receiving state for a node, and the second half means the probability of receiving an entire packet which a neighbor node sends. A receiver cannot receive a packet if the duration of the sender's transmit (Tx) mode is not completely comprised in that of the receiverfs receive (Rx) mode. The entire equation implies that if a node has a high charging rate, then it also has high reception probability. If $p_i = 0$, then it means that the node is unable to charge and therefore is unable to operate at all, whereas if $p_i = 1$, then it means that the node is poweblack by a stable source and is always ready to receive, such as the sink node. Node i sets the initial value of neighbor node j's reception probability p_j to 1 and updates p_j when node i receives a packet from node j. τ should be longer than one cycle (charge, receive and transmission), but it becomes increasingly harder to adapt to a change of environment as τ become larger; consequently, the optimal τ depends on the network environment, such as, the type of energy harvest device that node uses and its charging rate.

When node n_i broadcasts a packet once, the probability q_i that at least one node in S_i receives the packet for forwarding is expressed as:

$$q_i = 1 - \prod_{j \in S_i}(1-p_j). \quad (2)$$

According to Equation (2), if there are nodes with high reception probability (p_j) within range of node n_i or the number of nodes within range is large, then q_j is high. Similarly, assuming a sender (n_i) repeats transmitting a given packet a times, then the probability that at least one node in S_i receives the packet at least once, r_{ai}, is calculated as:

$$r_{ai} = 1-(1-q_i)^a.$$

In PRT, let Th denote the reliability threshold on q_i, and we assume that $r_{ai} \geq Th$ is reliable. Then, a sender calculates an optimal number of retransmission, denoted by a', which is the minimum value such that the probability $r_{a'i}$ remains above the reliability threshold, as follows:

$$\begin{aligned} r_{a'i} &\geq Th \\ 1-(1-q_i)^{a'} &\geq Th \\ (1-q_i)^{a'} &\leq 1-Th. \end{aligned}$$

Taking $\log_{(1-q_i)}$ on both sides,

$$a' \geq \log_{(1-q_i)}(1-Th). \quad (3)$$

After the sender calculates a', it then transmits a packet a' times. The Th value should be chosen based on the reception probability of the neighboring nodes. As Th increases, so does the optimal number of repeated transmissions a' and energy consumption. Therefore, it is expected that an optimal Th value exists for different settings.

B. PRT with Collision Consideration (PRT-CC)

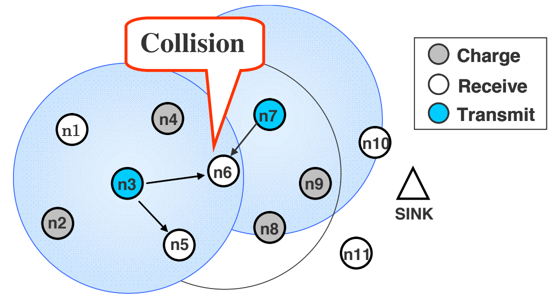


Fig. 5. The nodes n_7, n_8 and n_9 are hidden terminals of node n_3

In PRT, the reception probability p_i of node n_i is computed without considering collisions. Therefore, a' can be estimated to be smaller than the optimal value when collisions occur frequently. Then, in PRT-CC, each node computes the reception probability with collision consideration.

Figure 5 shows an example that node n_3 broadcasts a packet. Nodes in $N_3 = \{n_1, n_2, n_4, n_5, n_6\}$ cannot enter the transmit state because the nodes do carrier sensing before starting transmission as shown in Section III. However, nodes $\{n_7, n_8, n_9\}$ that are outside communication range of node n_3, cannot know that node n_3 is transmitting a packet. At this time, if a node in $N_6 \setminus N_3 \setminus \{n_3\}$ starts transmission, n_6 cannot receive the packet sent from n_3 correctly (i.e., the nodes $\{n_7, n_8, n_9\}$ are hidden terminals of node n_3).

To derive the reception probability with collision consideration, first, we introduce c_i to denote the probability that a node n_i is in the transmit state. Similar to the first half of the right side of Equation (1), given a node n_i and its measublack total charging/receiving/transmitting time, T_{chi}, T_{rxi} and T_{txi}, respectively, c_i is approximated as follows:

$$c_i = \frac{T_{txi}}{T_{chi}+T_{rxi}+T_{txi}}. \quad (4)$$

In PRT-CC, then, given a node n_i and its neighboring node n_j of n_i, the reception probability p'_{ji} that node n_i can receive a packet sent from n_i can be calculated as follows:

$$p'_{ji} = p_j \cdot \prod_{k \in N_j \setminus N_i \setminus \{n_i\}} (1 - c_k). \quad (5)$$

Equations (4) and (5) imply that if there is a node that transmits packets frequently in $N_6 \setminus N_3 \setminus \{n_3\}$, then the collision probability will increase, and the reception probability will decrease. Moreover, the probability that at least one node can receive the packet when node n_i broadcasts a packet once is expressed as Equation (6):

$$q'_i = 1 - \prod_{k \in S_i} (1 - p'_{ki}). \quad (6)$$

Computing a' by Equation (3) with q'_i instead of q_i, nodes can estimate the optimal number of times for retransmission with collision consideration. In order to ensure that a packet is transmitted a' times, PRT and PRT-CC use the duplicate detection technique of GR-DD, where if a node receives the same data packet multiple times, then the latter packets are discarded. In PRT-CC, more overhead is incurblack than PRT, due to the use of information of the neighboring nodes' transmission probability, in addition to the reception probability.

Note that, generally, packet corruption is mainly caused by collision. In this simulation, therefore, we assume that packet corruption occurs only if a collision occurblack. We will consider other reasons about corruption, such as bit error, in the future work.

C. Exchanging Packet Reception Probability

In both the PRT and PRT-CC protocols, a sender node uses its neighboring nodes' packet reception probability to compute the number of times for retransmission. Since each node computes its own packet reception probability, there needs to be a way for a node to obtain this information from its neighboring nodes.

One way of exchanging the information of probability among nodes is to schedule a time interval for information exchange. During this interval, all nodes update the probability of their neighbors simultaneously. However, in energy-harvesting WSNs, not all nodes have the adequate amount of charged energy to perform this exchange task at the scheduled time, making this scheme difficult to realize. Furthermore, time synchronization is difficult, if not impossible, to achieve when nodes run out of energy.

Another method is to piggyback the packet reception probability onto regular data packets. The advantages are that this does not require a special time interval dedicated to just information exchange, and it does not require the model of operation to be modified. Being a broadcast model means that multiple receivers can receive simultaneously, and each node just needs to update its local information. The overhead results in higher energy consumption and makes charging time longer. However, the additional data overhead requiblack is only the reception probability information. Depending on the

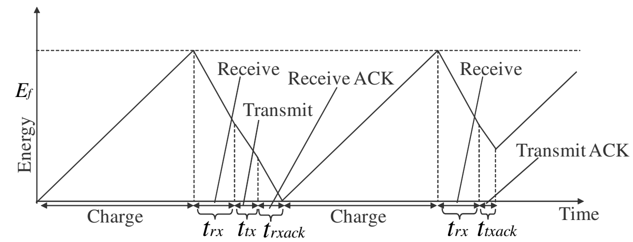


Fig. 6. Energy level over time for ACKnowledgement-based approach

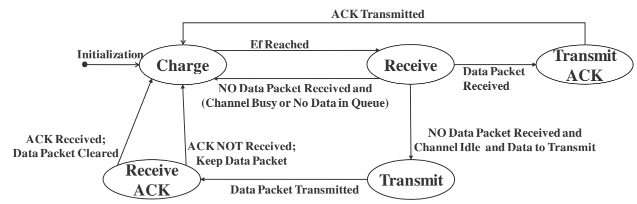


Fig. 7. FSM of node for ACKnowledgement-based approach

desiblack accuracy of the information of the probability, e.g. if we assume it is 8 bits, then PRT-CC requires 32 bits of extra information for the reception probability and collision probability.

D. Acknowledgement-based Retransmission (ACK)

Without the use of ACKs, nodes cannot detect packet loss [11] and have to rely on other mechanisms to improve reliability. While our proposed approach aims to offset the packet loss by retransmitting packets probabilistically instead of relying on ACKs, we also study the use of ACKs in WSNs poweblack by energy harvesting for comparison against our proposed schemes.

We model the nodes' operation when ACKs are used. Broadcast-based communication is used, as described in Section III, due to the unpblackictable energy source. The amount of stoblack energy on a node over time is shown in Fig. 6, and each node operates according to the finite state machine (FSM) shown in Fig. 7.

To transmit and receive the ACKs, more energy is needed, and therefore E_{fack}, the minimum amount of energy of this model, is larger than E_f. E_{fack} is calculated as:

$$E_{fack} = P_{rx} \cdot t_{rx} + P_{tx} \cdot t_{tx} + P_{rx} \cdot t_{rxack}.$$

t_{rxack} denotes the receiving time for ACK. After a node charges up to E_{fack}, it enters the **Receive** state to receive data from neighboring nodes. If the node has received data during the **Receive** state, the node transmits the ACK to the sender. The transmission time for the ACK, denoted by t_{txack}, is derived from

$$t_{txack} = \frac{s_{ack}}{\alpha}$$

where s_{ack} denotes the size of the ACK packet and the transmission rate is α. Otherwise, if there is a packet in the queue when the receive state ends and the channel is idle, then the node transmits packet. After the transmission, the

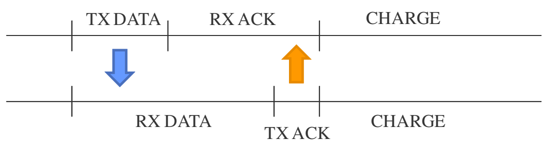


Fig. 8. Node receives data at the start of the **Receive** state

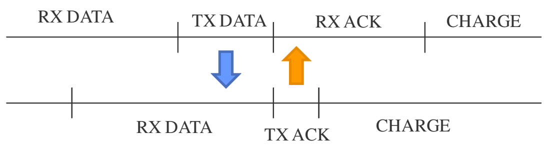


Fig. 9. Node receives data at the end of the **Receive** state

nodes moves to the **Receive ACK** state to wait for the ACK from the receiver. The receiving time for the ACK, t_{rxack}, must be sufficiently long to ensure that the sender can hear the ACK from the receiver. Figure 8 shows the case that a node receives data at the start of the receiving state and transmits the ACK to the sender. Conversely, Fig. 9 shows the case that a node receives data at the end of the receiving state. From these cases, we can see that $t_{rxack} = t_{rx} - t_{tx} + t_{txack}$ is sufficient to receive the ACK from the receivers in any case. Regardless whether a node receives the ACK or not, the node returns to the charge state until it charges up to E_{fack} again. Neighboring nodes cannot always receive data from the sender, and collision of ACK can occur if more than one node transmit ACKs at the same time. Consequently, a node will transmit the same data again in the next cycle, if the node did not receive the ACK from any neighboring nodes.

V. Evaluations

We evaluate our proposed protocols by simulation. In this simulation, n nodes are deployed randomly in the area as shown in Fig. 10. The area size is 500m $\times$ 500m. There is a sink node and it is located at the center of the area. This sink

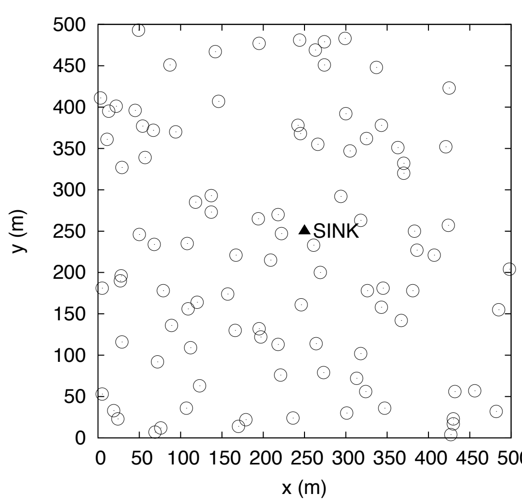


Fig. 10. Random Topology

TABLE I
Simulation Parameters

Simulation Time	1000 (s)
Transmission Power Consumption P_{tx}	76.2 (mW)
Receiving Power Consumption P_{rx}	83.1 (mW)
Transmission Power	-3.0 (dbm)
Receiving Power	-85.0 (dbm)
Transmission Antenna Gain	0 (dbi)
Receiving Antenna Gain	0 (dbi)
Frequency f	2.4 (GHz)
Data size	800 (bits)
Transmission Rate α	250 (kbps)

is powered by a steady power supply so that it can be always ready to receive. In the simulation, we assume that there is no packet loss within the communication range, but collision means no correct packet can be received. The communication range of node is computed by Friis transmission equation [15]. A communication range is about 125m where the decay factor (m) of the Friis transmission equation is 2. At all nodes, data packets are generated according to a Poisson distribution with an arrival rate of λ packet/s and it is set to 0.1 in the simulation. We assume that the original packet size s is 800 bits. To provide the necessary information for our protocols, the size of a packet for PRT is 808 bits including 8 bits of reception probability, and that for PRT-CC is 832 bits including 32 bits of reception probability and collision probability.

Since the charging rate of each node varies with time, we denote the average of charging rate by Ch and the variation of Ch by v. Given a node, the charging rate of the node is generated by the uniform distribution within $[Ch - v, Ch + v]$ for each instance of time. In the simulation, Ch is 10mW similar to [11] and v is fixed at a value of 2mW.

In PRT and PRT-CC, the reliability threshold Th is set to 0.9, the measurement period τ is set to ∞ because the functions using τ is only for the network where the average charging rate varies greatly. The transmission/receiving power, data rate, and various parameters of a node are based on those of the MICAz platform. The simulation parameters are shown in Table I.

The evaluation metrics are delivery ratio and delay. The delivery ratio is defined to be the unique packet count as received by the sink divided by the total generated packet count. The delay is the time from the instant a packet is generated till the instant it reaches the sink node. To evaluate the performance of the PRT and PRT-CC protocols, we compare our results with those for GR-DD and GR-DD-RT. We evaluate the proposed protocols, with different parameter values, namely, the number of nodes n, the average of charging rate Ch, the decay factor m, the arrival rate λ and the reliability threshold Th.

A. Number of Nodes (n) vs. Delivery Ratio

We evaluate the delivery ratio of each protocol by varying the number of nodes, as it affects the node density and delivery ratio. The results in Fig. 11 show that PRT, PRT-CC and ACK achieve a higher data delivery ratio than the GR-DD and GR-DD-RT in most cases.

As the number of nodes in the area increases, the delivery

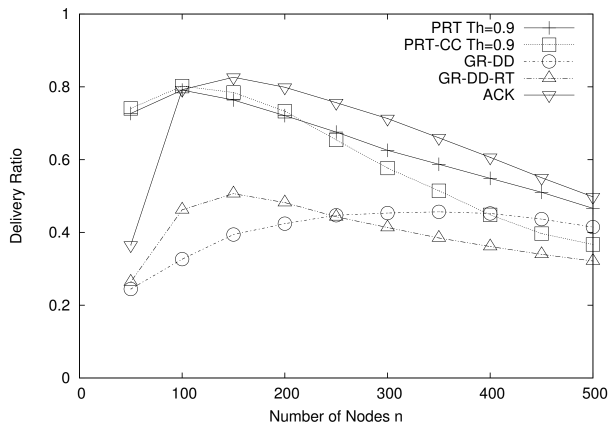


Fig. 11. Number of Nodes n vs. Delivery Ratio

ratio also increases for networks with 150 nodes or fewer. Since the number of nodes in the communication range of each node increases, the number of nodes which can receive a packet when a sender broadcasts it also increases. However, the delivery ratio of every protocol except GR-DD decreases with more than 150 nodes. While GR-DD transmits individual data only once, other protocols retransmit the same data and consume more energy. Therefore, nodes cannot transmit all data when large amount of data is generated by many nodes.

With the use of ACK, nodes can detect packet loss and retransmit data to their neighboring nodes, thus, achieving the highest delivery ratio among the protocols when the number of nodes exceeds 100. In the case of 50 nodes, the probability that the neighboring nodes receive data from a sender is low because the number of neighboring nodes is too small. Additional energy needed to transmit and receive the ACKs reduces the chance to receive packets and results in low delivery ratio.

In PRT and PRT-CC, each node can adapt to the changes in the node density by computing the optimal number of times to retransmit a packet depending on the number of nodes within its communication range. The results show slightly better delivery ratio in PRT-CC compared to PRT only when the number of nodes is small. This is because the larger packet size of PRT-CC adversely affects its delivery ratio, which becomes worse than PRT, when the number of nodes is large. The number of retransmissions by PRT-CC is larger than that of PRT, so nodes cannot send out much of the data in the queue. On the other hand, PRT is effective even though its operation is simple.

B. Charging Rate (Ch) vs. Delivery Ratio

The charging rate of a node, which is derived from the energy harvesting rate, is also a critical factor which affects the system performance directly. According to [16], the charging rate on a 10cm^2 energy harvesting material (which is about the same size as a wireless sensor mote) is from 0.032mW (indoor) to 37mW (direct sunlight) when energy is harvested from solar or 5mW (piezoelectric) when energy is harvested from vibration. Besides the default rate of 10mW, we also evaluate the protocols under different charging rates and the simulation results are shown in Fig. 12.

Every protocol achieves the higher delivery ratio with the higher average charging rate *Ch* when the number of nodes in the area is small (Fig. 12(a)). As the charging rate increases, nodes can charge quickly and can operate at higher duty cycle. Nodes have more opportunities for receiving/transmitting packets and can deliver more packets to the sink. When the charging rate is small, PRT and PRT-CC achieve higher delivery ratio than the other protocols. The delivery ratio of ACK is quite low with small *Ch* values and it increases rapidly as *Ch* increases. Nodes must stay in the **Charge** state longer to transmit/receive ACK when the charging rate is small, and this degrades the delivery ratio. The better performance of ACK at higher charging rates is expected because the energy profile becomes more similar to that of conventional WSNs with consistent power sources.

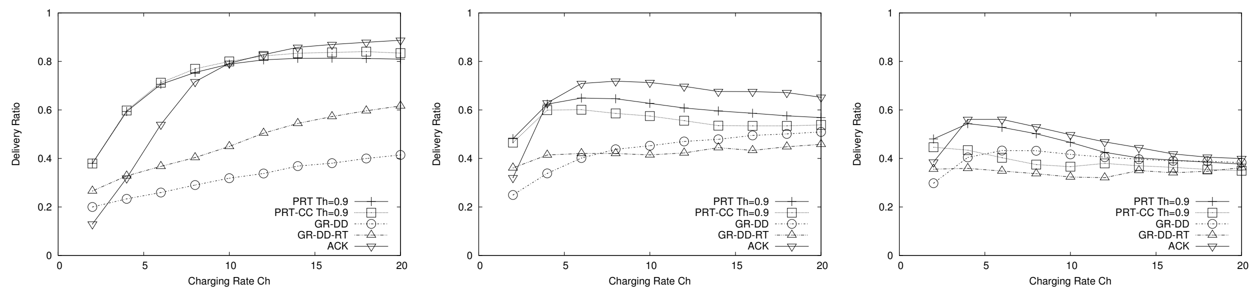
When more nodes are deployed in the area (Fig. 12(b) and (c)), the delivery ratio is low and shows small variations regardless of the charging rate because of large volume of data in the network generated by more nodes. However, the results show that PRT, PRT-CC and ACK provide superior performance consistently compared to GR-DD and GR-DD-RT in most cases.

While the charging rate is expected to improve as the energy harvesting technology advances, various unpredictable environmental factors can result in low energy harvesting rates which protocols must be designed to handle.

C. Decay Factor (m) vs. Delivery Ratio

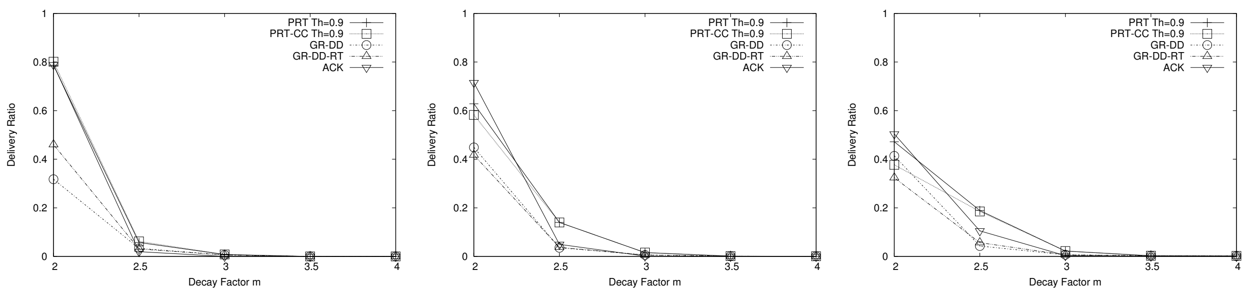
In PRT and PRT-CC, each node uses the information of neighboring nodes within its communication range. The communication range of a node depends on the quality of the wireless link. At the same transmission power, a node can transmit a signal further in obstacle-free space than one with obstacles. In the Friis transmission equation, the decay factor m represents the path quality, where larger m denotes worse condition for signal transmission. m is measured empirically, $m = 2$ refers to free (outdoor) space and $m = 2.5$ means the indoor space while $m \geq 3$ means the environment where it is not good for wireless communication, approximately. In this aspect of the evaluation, we vary the decay factor m of the Friis transmission equation to change the communication range of nodes, and therefore, the number of nodes within the communication range of an arbitrary node.

Figure 13 shows the results of the packet delivery ratio of each of the protocols for different values of m. At higher values of m, with transmission power unchanged, the effective communication range is reduced. As expected, every protocol shows low delivery ratio with large m. In the case with m larger than 3, nodes cannot deliver most of the packets in every protocol. At shorter ranges, the number of neighbor nodes become fewer, and consequently, the delivery ratio decreases. To overcome this problem, nodes should use higher power to transmit packets at the cost of longer charging times.



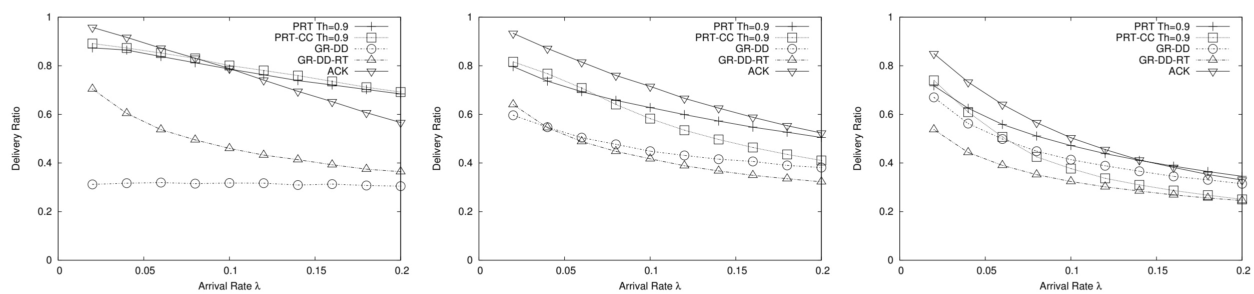
(a) $n = 100$ (b) $n = 300$ (c) $n = 500$

Fig. 12. Charging Rate *Ch* vs. Delivery Ratio



(a) $n = 100$ (b) $n = 300$ (c) $n = 500$

Fig. 13. Decay Factor *m* vs. Delivery Ratio



(a) $n = 100$ (b) $n = 300$ (c) $n = 500$

Fig. 14. Arrival Rate λ vs. Delivery Ratio

D. Arrival rate (λ) vs. Delivery Ratio

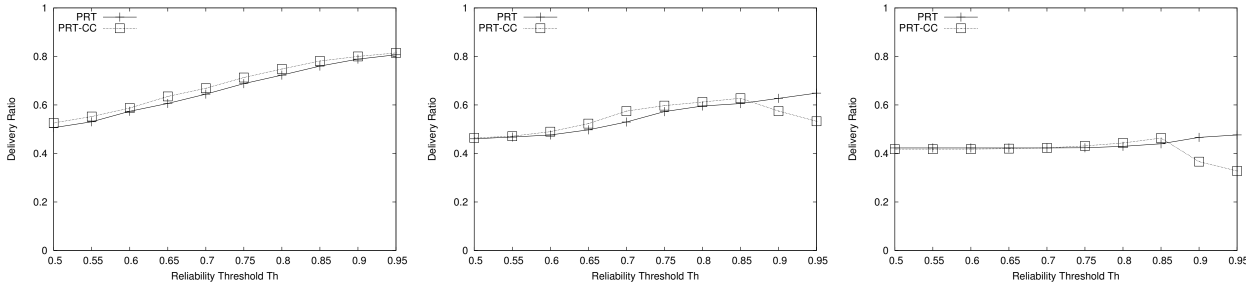
The data generation process of sensors depends on applications. For example, the sensors used to monitor temperature and humidity in a farm generate the data periodically, while an event-monitoring system generate the data when it detects the occurrence of events. Furthermore, the frequency of the data generation differs from one application to another. In the performance evaluation, we use a Poisson distribution with an arrival rate of λ packet/s to model the data generation process and evaluate the delivery ratio under different values of λ.

The results in Fig. 14 show that the delivery ratio decreases as the arrival rate increases in every protocol. Nodes powered by energy harvesting are constrained by the limited and unpredictable energy supply, and hence the amount of data that the nodes can transmit is also limited. This is particularly obvious in the protocols that retransmit the same packet more times, namely, PRT-CC and ACK, where the delivery ratio decreases substantially when a large volume of data is generated at higher data arrival rates.

E. Reliability Threshold (Th) vs. Delivery Ratio

In PRT and PRT-CC, we are able to control the desired performance by assigning different values to the reliability threshold *Th*. Figure 15 shows the delivery ratio for different values of *Th*. When *Th* is set to a large value, nodes retransmit the same packet many times to achieve high reliability. However, the delivery ratio of PRT-CC drops when large *Th* values (i.e. $Th \geq 0.85$) are used in dense networks ($n \geq 300$). In the case where large volume of data are generated at a node (high data arrival rates) or low charging rates are experienced, if *Th* is set to a large value, a node cannot transmit all its packets fast enough, resulting in low delivery ratio. Therefore, *Th* should be appropriately chosen depending on the node's condition.

Moreover, *Th* also depends on the application. If the system is used in critical applications like disaster information



(a) $n = 100$ (b) $n = 300$ (c) $n = 500$

Fig. 15. Reliability Threshold *Th* vs. Delivery Ratio

networks and structural health monitoring [17], *Th* should be set to a large value to provide high accuracy and reliability. On the other hand, a large *Th* is not needed when the system is used in applications high reliability is not critical, such as, environmental monitoring [18].

F. Delay

Some critical realtime sensor network applications need to collect data quickly. Therefore, we evaluate the delay with different number of nodes (n) in the network, the average charging rate (Ch) and the arrival rate (λ), as shown in Figs. 16, 17 and 18, respectively. Figures 16 and 18 show the delay in PRT-CC and ACK are larger than other protocols when a large volume of data are generated. This is because some packets reach the sink node after multiple retransmissions by these protocols.

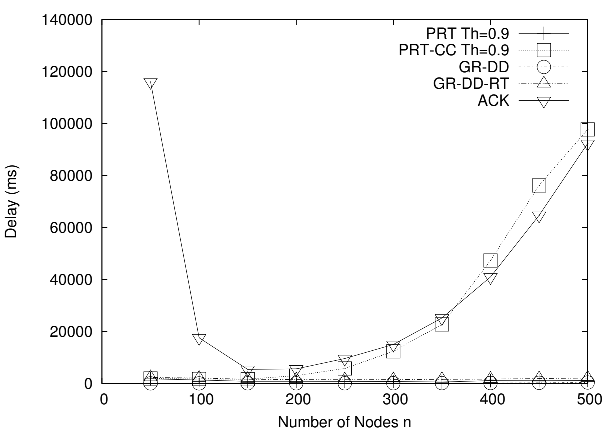


Fig. 16. Number of Nodes n vs. Delay

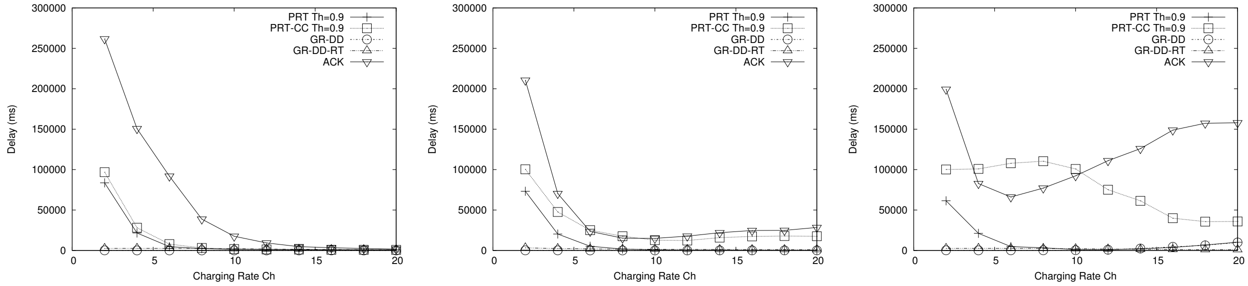
As the average charging rate Ch increases, the delay decreases in almost every case due to the shorter charging time (Fig. 17). However, the delays in PRT, PRT-CC and ACK are larger than GR-DD and GR-DD-RT especially when Ch is small, for example, when the sun is obscured by clouds. The reason for the lower delays in GR-DD and GR-DD-RT can be attributed to their lower delivery ratios, since many packets from nodes far away from the sink cannot reach the sink node and the computation of the delay is based on the few packets that actually reached the sink. The results also show that the ACK takes a longer time to deliver the packets to the sink because the nodes take a long time to charge as more energy is needed to transmit and receive the ACKs. Moreover, the delay when ACKs are used increases despite the higher charging rate ($Ch > 10$) with the 500-node network (Fig. 17(c)). When the number of nodes in the network is large and the charging rate is high, the number of nodes that become active concurrently also increases and that makes the network more congested, resulting in more packet collisions. When ACKs are used, nodes retransmit the same data until the ACKs are successfully received while other protocols have a limit on the number of retransmissions. Therefore, the number of packets that reach the sink after many retransmissions can be large, and this increases the overall delay.

G. Overhead

The strategy to utilize energy with energy harvesting devices for sensor nodes is different from that with conventional battery. The remaining energy level of an energy harvesting device can increase by charged from ambient energy sources whereas that of a conventional battery is monotonically decreasing. The energy capacity of such energy harvesting devices is generally too small to keep it for later, so it is reasonable to use energy from moment to moment. PRT is based on GR-DD, in other words PRT without retransmission is same as GR-DD. As shown in Figs. 12 to 15, PRT performs better than GR-DD, and it indicates that the traffic due to retransmission does not unnecessary. The possible side effects of too many retransmission are follows: the reception probability is getting low because T_{txi} in Equation (1) is getting large; some packets can be dropped due to overflow at the TX queue with high node density and high arrival rate situations. These effects can be avoidable if appropriate Th is chosen.

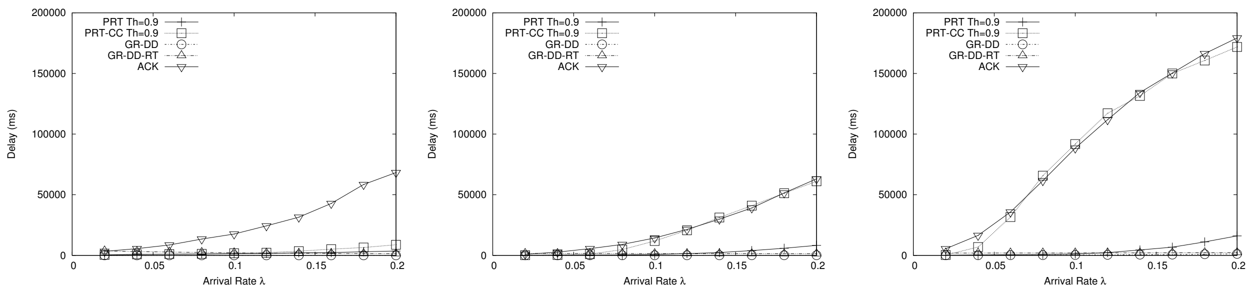
In PRT-CC, in order to comprehend the presence of hidden terminals, each node disseminates the list of its neighbor nodes piggybacking the list in the header of a data packet in PRT-CC. We have evaluated the impact of this overhead when the header size is 32bit, 64bit, 128bit and 256bit whereas the size of payload of a data packet is 800bit. The size of the list depends on the node density and the size of communication range.

Figure 19 shows the impact of the packet header size for PRT-CC. The overhead from this size is small when the



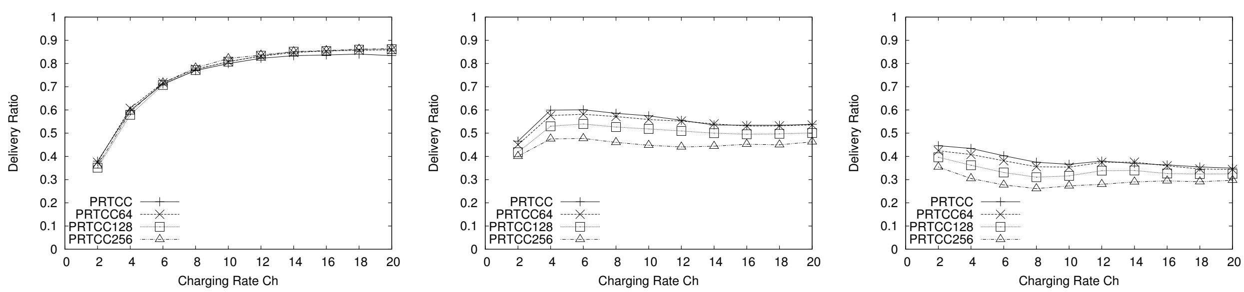
(a) $n = 100$ (b) $n = 300$ (c) $n = 500$

Fig. 17. Charging Rate *Ch* vs. Delay



(a) $n = 100$ (b) $n = 300$ (c) $n = 500$

Fig. 18. Arrival Rate λ vs. Delay



(a) $n = 100$ (b) $n = 300$ (c) $n = 500$

Fig. 19. Overhead of Packet Header Size (Charging Rate *Ch* vs. Delivery Ratio)

number of nodes is small.

H. Different Energy Harvesting Profile

Realistically, each sensor node has a different energy harvesting profile because the amount of harvested energy depends on the environment where nodes deployed and the device of nodes. With solar energy harvesting, nodes deployed in a sunny location have higher charging rates while nodes in the shade have lower charging rates. Therefore, nodes within a locality tend to exhibit similarity in their energy harvesting profile.

To evaluate in effect of differing harvesting rates, the target area is divided into 16 subareas as shown in Fig. 20. We choose four subareas and give the lower charging rate (2mW) to the nodes in these subareas while other nodes have higher charging rate (10mW). Nodes in the same subarea have the same average charging rate due to the correlation between the energy harvesting profile of nodes. Figures 21 and 22 illustrate the results for the scenario where four subareas around the sink are chosen. Figures 23 and 24 illustrate the results for the scenario where four subareas are chosen from the 12 subareas away from the sink randomly.

Results in Fig. 21 and 22 show the lower delivery ratio compared to ones in Figs. 11 and 14. Generated data are relayed to the sink by multihop, so the amount of data that node need to transmit increases as the distance between the node and the sink decreases. Although the nodes near the sink have many data received from other nodes, the nodes do not have enough energy to send out all data, then the nodes become the bottleneck in the network and degrade the delivery ratio in this scenario. Especially, the ACK-based approach is heavily affected by the lower charging rate in this scenario. In Fig. 21, the delivery ratio when ACKs are used decreases rapidly as more data are generated (larger λ) because the

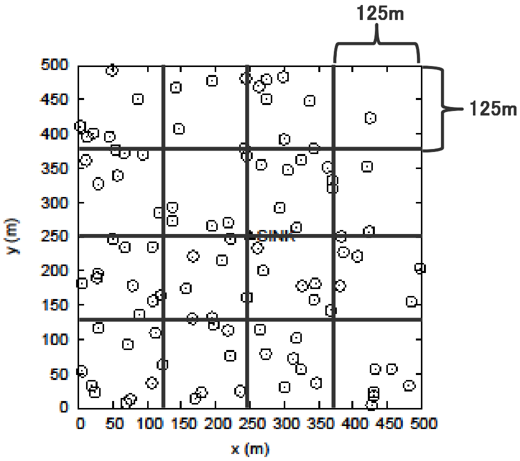


Fig. 20. Area divided into Subareas

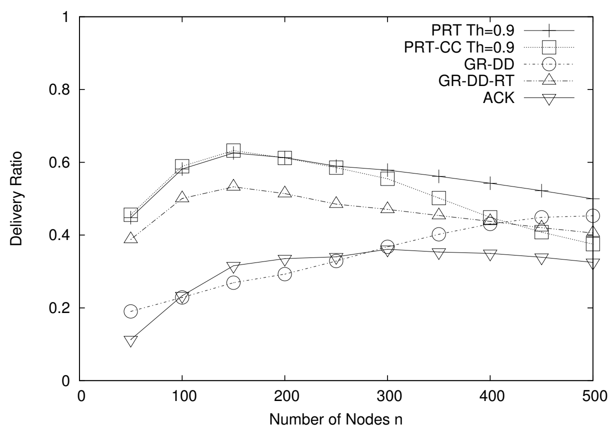


Fig. 21. Number of Nodes *n* vs. Delivery Ratio (Different energy harvesting profile for nodes around sink)

node needs more energy transmit/receive ACKs than other protocols.

Conversely, Fig. 23 and 24 show little difference from Fig. 11 and 14. Nodes away from the sink do not need to relay many data from other nodes, so the energy harvested with the lower charging rate are enough to send out the data even when some nodes have the lower charging rate. From these results, it is better to place the sink node in the sunny area where nodes can harvest more energy.

VI. Conclusions

In this paper, we discussed data collection for energy harvesting wireless sensor networks and showed that frequent packet loss as an intrinsic problem. In order to overcome the problem and to improve data collection efficiency, we proposed two data collection protocols named Probabilistic ReTransmission (PRT) and Probabilistic ReTransmission with Collision Consideration (PRT-CC), where a sender node calculates the probability of receiving a packet based on its own active intervals. We also considered ACKnowledgement-based approach and model the energy harvesting nodes' operation when ACKs are used. We evaluated PRT, PRT-CC and ACK-based approach by simulation, and the results show that PRT, PRT-CC and ACK-based approach achieve higher delivery ratio than the previous work (GR-DD and GR-DD-RT). For the future work, we are considering mathematical analysis of the models and protocols and how to determine the optimal Th and τ in PRT and PRT-CC.

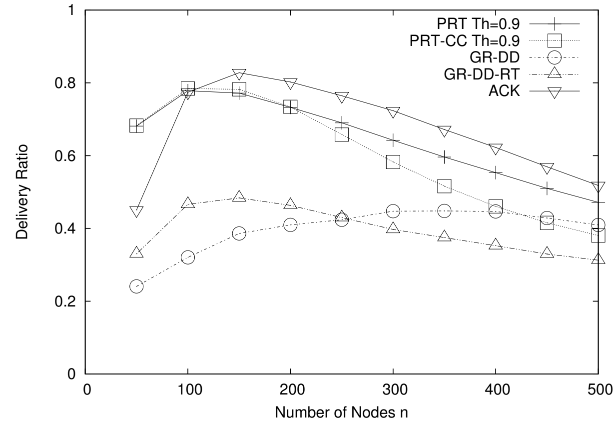


Fig. 23. Number of Nodes *n* vs. Delivery Ratio (Different energy harvesting profile for nodes away from sink)

References

[1] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proc. of the Hawaii Int'l Conf. on System Sciences*, 2000, pp. 1–10.

[2] W. Ye, J. Heidemann, and D. Estrin, "An energy-efficent mac protocol for wireless sensor networks," in *Proc. of the IEEE Computer and Communications Societies(INFOCOM)*, 2002, pp. 1567–1576.

[3] A. Kansal and M. Srivastava, "An environmental energy harvesting framework for sensor networks," in *Proc. of the 2003 International Symposium on Low Power Electronics and Design*, 2003, pp. 481–486.

[4] C. Park and P. Chou, "AmbiMax: Autonomous energy harvesting platform for multi-supply wireless sensor nodes," in *Proc. of SECON*, 2006, pp. 168–177.

[5] K. Lin, J. Hsu, S. Zahedi, D. C. Lee, J. Friedman, A. Kansal, V. Raghunathan, and M. B. Srivastava, "Heliomote: Enabling long-lived sensor networks through solar energy harvesting," in *Proc. of ACM Sensys*, November 2-4 2005.

[6] H.-P. Tan, P. Lee, W. K. G. Seah, and Z. Eu, "Impact of Power Control in Wireless Sensor Networks Powered by Ambient Energy Harvesting (WSN-HEAP) for Railroad Health Monitoring," in *Proc. of the Second International Workshop on Applications of Ad hoc and Sensor Networks (AASNET09)*, Bradford, UK, May 26-29 2009, pp. 804–809.

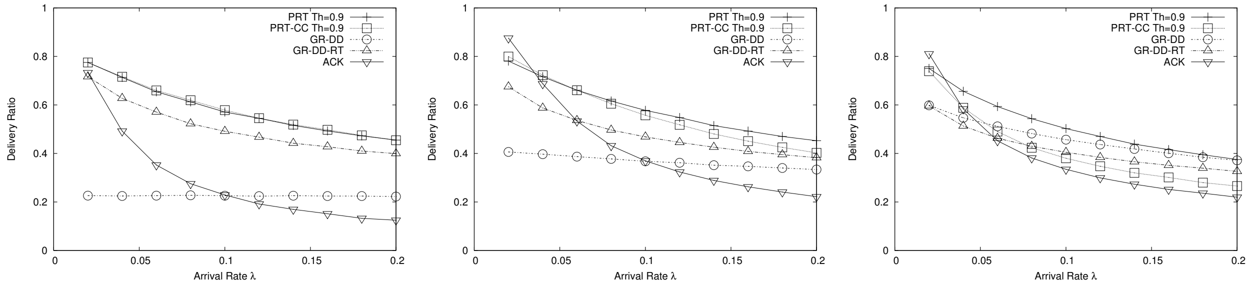
[7] M. Rahimi, H. Shah, G. Sukhatme, and J. Heideman, "Studying the feasibility of energy harvesting in a mobile sensor network," in *Proc. of IEEE Int'l Conf. Robotics and Automation*, 2003, pp. 19–24.

[8] A. Kansal, S. Z. J. Hsu, and M. B. Srivastava, "Power management in energy harvesting sensor networks," *ACM Transactions on Embedded Computing Systems*, 2006.

[9] K.-W. Fan, Z. Zheng, and P. Sinha, "Steady and fair rate allocation for rechargeable sensors in perpetual sensor networks," in *Proc. of the 6th ACM conference on Embedded network sensor systems*, 2008, pp. 239–252.

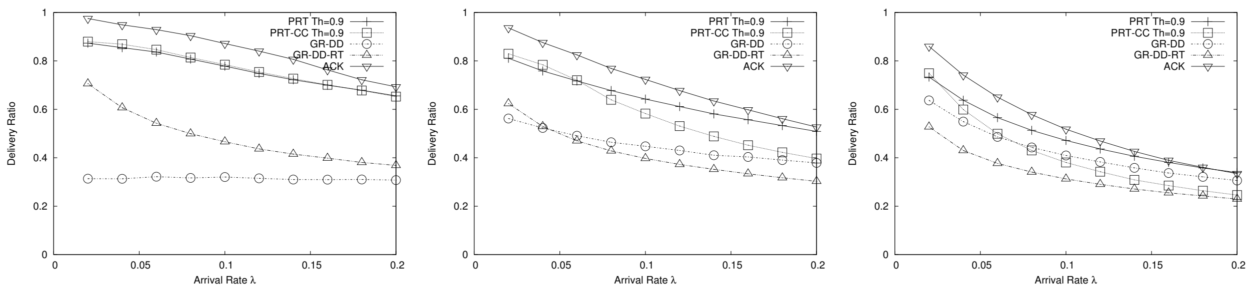
[10] V. Joseph, V. Sharma, and U. Mukherji, "Optimal sleep-wake policies for an energy harvesting sensor node," *PEDIATRICS*, vol. 115, no. 1, pp. 250–256, 2005.

[11] Z. A. Eu, H.-P. Tan, and W. K. G. Seah, "Routing and Relay Node Placement in Wireless Sensor Networks Powered by Ambient Energy Harvesting," in *Proc. of the IEEE Wireless Communications and Networking Conference (WCNC)*, Budapest, Hungary, Apr 5-8 2009, pp. 5–8.



(a) $n = 100$ (b) $n = 300$ (c) $n = 500$

Fig. 22. Arrival Rate λ vs. Delivery Ratio (Different energy harvesting profile for nodes around sink)



(a) $n = 100$ (b) $n = 300$ (c) $n = 500$

Fig. 24. Arrival Rate λ vs. Delivery Ratio (Different energy harvesting profile for nodes away from sink)

[12] ——, “Design and performance analysis of mac schemes for wireless sensor networks powered by ambient energy harvesting,” *Ad Hoc Networks*, vol. 9, no. 3, pp. 300–323, 2011.

[13] K. Tutuncuoglu and A. Yener, “Short-term throughput maximization for battery limited energy harvesting nodes,” in *Proc. of the IEEE International Conference on Communications (ICC)*, June 5-9 2011, pp. 1–5.

[14] I. Schweizer, N. Fleischhacker, M. Mühlhäuser, and T. Strufe, “Sdf - solar-aware distributed flow in wireless sensor networks,” in *Proc. of the IEEE Conference on Local Computer Networks (LCN)*, October 4-7 2011, pp. 386–394.

[15] H. Friis, “A note on a simple transmission formula,” *Proceedings of the IRE*, vol. 34, no. 5, pp. 254–256, May 1946.

[16] W. K. G. Seah, Z. A. Eu, and H.-P. Tan, “Wireless Sensor Networks Powered by Ambient Energy Harvesting (WSN-HEAP) – Survey and Challenges,” in *Proc. of the 1st International Conference on Wireless Wireless Communications, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology (Wireless VITAE)*, May 17-20 2009.

[17] C. Park, Q. Xie, P. Chou, and M. Shinozuka, “DuraNode: wireless networked sensor for structural health monitoring,” in *Proc. of the IEEE International Conference on Sensors 2005*, 2005, pp. 277–280.

[18] A. Mainwaring, J. Polastre, R. Szewczyk, D. Culler, and J. Anderson, “Wireless sensor networks for habitat monitoring,” in *ACM International Workshop on Wireless Sensor Networks and Applications*, 2002.