Multi-Attribute Data Fusion for Energy Equilibrium Routing in Wireless Sensor Networks

Kai Lin[1], Lei Wang[2], Keqiu Li[1] and Lei Shu[3]
[1]School of Electronic and Information Engineering, Dalian University of Technology, Dalian, China
[e-mail: {link, keqiu}@dlut.edu.cn]
[2]School of Software Engineering, Dalian University of Technology, Dalian, China
[e-mail: lei.wang@ieee.org]
[3]Nishio Lab., Department of Multimedia Engineering, Graduate School of Information Science and Technology, Osaka University, Japan
[e-mail: lei.shu@ieee.org]
*Corresponding author: Kai Lin

Abstract

Data fusion is an attractive technology because it allows various trade-offs related to performance metrics, e.g., energy, latency, accuracy, fault-tolerance and security in wireless sensor networks (WSNs). Under a complicated environment, each sensor node must be equipped with more than one type of sensor module to monitor multi-targets, so that the complexity for the fusion process is increased due to the existence of various physical attributes. In this paper, we first investigate the process and performance of multi-attribute fusion in data gathering of WSNs, and then propose a self-adaptive threshold method to balance the different change rates of each attributive data. Furthermore, we present a method to measure the energy-conservation efficiency of multi-attribute fusion. Based on our proposed methods, we design a novel energy equilibrium routing method for WSNs, viz., multi-attribute fusion tree (MAFT). Simulation results demonstrate that MAFT achieves very good performance in terms of the network lifetime.

1. Introduction

Wireless sensor networks (WSNs) comprise integrated technologies in combination with a multidisciplinary approach, such as sensing, micromachining systems, embedded computing and network communication. The appearance of WSNs is not only helpful for realizing ubiquitous computing, but also greatly improves the interaction between the human and physical world [1][2]. Currently, most of the crucial technologies in WSNs are poorly developed while the great challenge is the limitation of energy supply, computing capacity and network communication bandwidth. Therefore, one of the essential research problems is how to effectively utilize limited resources to achieve good data gathering performance.

In WSNs, the high deployment density of sensor nodes can result in content redundancy of sensory data packets transmitted in data gathering. If all redundant packets are transmitted, energy and bandwidth will be heavily wasted. For the sake of improving the resource utilization, data fusion is employed in WSNs [3][4][5]. The main idea of data fusion is to aggregate the multi-data into more efficient data. As data fusion can reduce the abundant data, unnecessary energy consumption is avoidable. Besides, data fusion is helpful to acquire accurate information, which can overcome the limitations of low accuracy sensors. Additionally, data fusion can reduce communication conflict. For these reasons, data fusion has become an indispensable technology for WSNs.

With the development of WSNs, data fusion is starting to be used in combination with other technologies and especially routing. Routing is not only responsible for data gathering, but has to provide the foundation for data fusion, time synchronization and target location [6][7]. The performance criterion of a good routing method is whether it can remove the unnecessary links and utilize the energy efficiently. Our research work in this paper focuses on how to combine the routing method and data fusion mechanism, and the design of a high efficiency fusion-driven routing method. To avoid energy holes caused by excessive energy consumption of partial sensor nodes, the establishment of the routing method should guarantee energy equilibrium. Based on energy equilibrium, it should aim for maximal reduction in energy consumption and prolong the lifetime of the whole network. Therefore, it is desirable to design a routing method which can guarantee energy equilibrium and fully utilize the energy-conservation efficiency of data fusion. Considering the diversity of WSNs, our research aims at complicated multi-attribute fusion. The main contributions of this paper are summarized as follows:

- We discuss the data fusion process in both homogenous and heterogeneous WSNs, and analyze the negative effect of a multi-attribute change rate difference on the fusion process. Based on the analysis, we propose a self-adapting threshold to solve this problem. This method can balance the different change rates of each attributive data.

- We analyze the energy-conservation efficiency of multi-attribute fusion in data gathering. For multi-attribute fusion, the correlations of packets required for fusion is much lower than those of single-attribute fusion. In this case, the computing cost for the complicated fusion process cannot be simply neglected. We consider both the effect of multi-attributes and the cost of data fusion in our research. Then we present a method to measure the energy-conservation efficiency of multi-attribute fusion.

- In combination with multi-attribute fusion and routing, we propose a new energy equilibrium routing scheme, viz., multi-attribute fusion routing (MAFT). Energy

equilibrium and conservation are both considered in MAFT. As well as guaranteeing some degree of energy equilibrium, the energy-conservation efficiency of multi-attribute data fusion can be maximized in MAFT.

- We perform extensive simulation experiments to evaluate MAFT by several performance criteria. The results show that MAFT achieves high efficiency performance as well as prolonging the lifetime of WSNs.

The rest of this paper is organized as follows. Section 2 presents some related works and section 3 introduces the system model and problem statements. In Section 4, we analyze multi-attribute fusion and propose a self-adaptive threshold. In Section 5, we present a method to measure the energy-conservation efficiency of multi-attribute fusion. In Section 6, we design a novel energy equilibrium routing, viz., multi-attribute fusion tree (MAFT). In section 7, the performance evaluation of MAFT by simulation is demonstrated. Finally, we summarize our work and conclude the paper in section 8.

2. Related Work

To date, researches have shown that data fusion (or aggregation) in WSNs may produce various trade-offs among some network related performance metrics such as energy, latency, accuracy, fault-tolerance and security. To date, it has become a crucial technology in WSNs, as many valuable research results have been obtained. It is attracting widespread attention as a means for combining data fusion with other technologies, such as fusion-driven routing.

In [8], Krishnamachari et al. investigated the impact of data aggregation on these networking metrics by surveying the existing data aggregation protocols in WSNs. In [9], Intanagonwiwat et al. proposed a novel approach that adjusted aggregation points in order to increase the amount of path sharing, and reduced energy consumption. The results suggested that greedy aggregation could achieve up to 45% energy savings over opportunistic aggregation in high-density networks, without adversely impacting latency or robustness.

In [10], Pattem et al. proposed several data aggregation techniques to study the performance of various data aggregation schemes across the range of spatial correlations. The analysis and simulations revealed that the characteristics of optimal routing with compression did depend on the level of correlation. Specially, there existed a practical static clustering scheme which could provide near-optimal performance for a wide range of spatial correlations.

In [11], Yu et al. employed the data aggregation tree to extract the packet flow. In [12], Goel et al. proposed a hierarchical matching algorithm, which resulted in an aggregation tree with simultaneous logarithmic approximation for all concave aggregation functions. In this model, each node can theoretically obtain the joint entropy of its sub-tree to achieve the maximal aggregation ratio. In [13], Cristescu et al. proved that the minimum-energy data gathering problem is NP-complete, by applying the reduction set-cover problem, and they claimed that the optimal result is an approximation based on a combination between the SPT and TSP.

In [14], Luo et al. proposed a routing algorithm called minimum fusion Steiner tree (MFST) for data gathering with aggregation in wireless sensor networks. MFST not only optimized over the data transmission cost, but also incorporated the cost for data fusion, which could be significant for emerging sensor networks with vectorial data and security requirements. In [15], they further found that fusion costs were comparable to those of communications for certain applications. Motivated by the limitations of MFST, they designed a novel routing algorithm,

called adaptive fusion Steiner tree (AFST) for energy efficient data gathering in sensor networks.

In [16], Richkenbach et al. proposed an optimal algorithm, MEGA, for foreign-coding and an approximation algorithm, LEGA, for self-coding. In MEGA, each node sent raw data to its encoding point using directed minimum spanning tree (DMST), and encoded data was then transmitted to the sink through SPT. In [17], LEGA uses shallow light tree (SLT) as the data gathering topology.

In [18], Luo et al. developed an online algorithm capable of dynamically adjusting the route structure when sensor nodes joined or left the network. Furthermore, by ensuring that such reconstructions were only performed locally and maximally preserving existing routing structure, the online algorithm could be readily implemented in real networks in a distributed manner, promised extremely small performance deviations from the offline version and outperformed other routing schemes with static aggregation decisions.

In [19], Anandkumar et al. presented a novel formulation for optimal sensor selection and in-network fusion for distributed inference known as the prize-collecting data fusion (PCDF) in terms of the optimal trade-off between the costs of aggregating the selected set of sensor measurements and the resulting inference performance at the fusion center. PCDF is then analyzed under a correlation model specified by a Markov random field (MRF) with a given dependency graph. For a special class of dependency graphs, a constrained version of PCDF reduces to the prize-collecting Steiner tree on an augmented graph. In this case, an approximation algorithm is given with an approximation ratio that depends only on the number of profitable cliques in the dependency graph.

In [20], Xing et al. bridged the gap between sensing coverage and the stochastic nature of sensing. The scaling laws were derived between coverage, network density and signal-to-noise ratio.

The weakness of all the above research works is that they do not consider the differences of attributes in packets that required fusion. However, these differences have direct effects on fusion results. Our research work is focusing on multi-attribute data fusion, which includes correctly measuring the effect of fusion on networks and combining it with routing.

3. System Models and Problem Statement

3.1 Network Model

We assume that all the nodes are uniformly distributed in a circular area A of radius R. Only one sink node is located at the center of area A. All the nodes have the same initial energy budget. In data gathering, the maximum communication distance is also the same for all nodes. Each node has a unique ID number and knowledge of its geographical location.

Without loss of generality, we make the following assumptions in this paper:
- All the sensor nodes and the sink node remain stationary after deployment.
- Except for the sink node, all the sensor nodes are isomorphic with the same initial energy, computation capacity and data fusion capacity. The sink node has no limitations of energy and computation capacity.
- Based on the distance to the receiver, the sensor nodes can adjust the transmission power to save energy consumption.
- When the sensor nodes have no tasks, they can switch to the sleeping state to save energy consumption.

3.2 Fusion Model

In a similar manner to [14][15], a data fusion model is employed in our research, where the sensor nodes are required to send their data constantly. In this model, a node v needs to receive the data sent from node u, which is denoted as $w(u)$. The total data amount after the fusion process at node v is expressed as:

$$w(v) = \max(\widetilde{w}(v), w(u)) + \min(\widetilde{w}(v), w(u))(1 - \rho) \tag{1}$$

Where $\widetilde{w}(v)$ represents the data amount generated by node v and ρ represents the data correlation between node u and v, which is affected by the data attributes.

3.3 Energy Model

We assume that all the nodes have the same initial energy E_0, while only the sink node has no energy limitations. In a similar manner to [21], the energy consumption of transmitting one-bit data over distance d is $e_t(d) = \varepsilon_{elec} + \varepsilon_{amp} \cdot d^k$, where ε_{elec} and ε_{amp} are the energy consumptions of the transmitter electronics and transmitting amplifier, respectively, and k ($k \geq 2$) is the propagation loss exponent. Consequently, the energy dissipation in receiving one-bit data is $e_r = \varepsilon_{elec}$. The data fusion process can introduce extra energy consumption, which is represented by e_f. Specially, the fusion cost will be increased by using encryption and other security mechanisms.

3.4 Problem Statement

Under a complex environment, it is necessary for WSNs to monitor multi-targets. To achieve this purpose, each sensor node needs to be equipped with more than one kind of senor, such as magnetometer sensor, thermal sensor, humidity sensor, pressure sensor, light sensor, sonic sensor etc. These sensors can monitor the environment separately or cooperatively. As a result, the generated packets contain multi-attribute data. The packet structure of the physical layer is shown in **Fig. 1**. When the data attributes in one packet are different, some of them could have a high ratio of changes while others might remain stable. When these packets are extracted or synthesized, the different change rates of the attributes makes the fusion process more complicated. Additionally, although data fusion can reduce redundant data and hence curtail network load, the energy consumption of the fusion process needs to be considered.
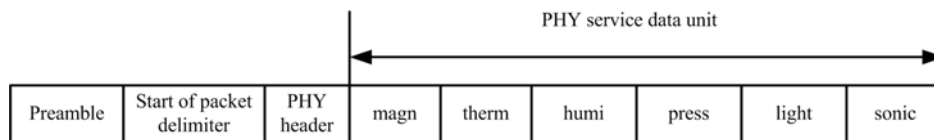

Fig. 1. Packet structure of PHY layer

For ease of exposition, we define the following two terms:

Definition 3.1. Energy equilibrium means that nodes in the network use up their energy simultaneously.

Definition 3.2. Change rate of data means the change speed of data content per unit time.

Now we begin to formulate the problem. For a given data-gathering sensor network, including the source nodes set S and sink node, each source node needs to send its data to the sink node. Due to communication capacity limitations of sensor nodes, the farthest source nodes from the sink node need m hops to send the data to the sink node. The sensor node set S_l represents nodes with l hops, where $\sum_{l=(1,2,\cdots,m)} S_l = S$ and S_0 is the sink node. Each source node belonging to S_l needs to select one forwarding node from S_{l-1} within its communication range. Our object is to design a routing protocol that can balance and minimize the energy consumption while delivering data from all source nodes in S to the sink node. This problem can be formulated as follows:

$$\begin{cases} \min \sum_{u \in S_l} \sum_{v \in S_{l-1}} e_{uv} \cdot x_{uv}, \quad l = (1,2,3,\ldots,m) \\ s.t. \quad for \ \forall u \in S_l, \ \sum_{v \in F_u} x_{uv} = 1 \\ E_v \geq E_k, \ u,k \in F_u \end{cases} \quad (2)$$

Where x_{uv} represents whether a connection exists between node u and v. When node v is the forwarding node of node u, then $x_{uv} = 1$, otherwise $x_{uv} = 0$. e_{uv} represents the energy consumption on the edge from node u to v comprising three components: node u transmitting data, node v receiving data and fusing data. It can be denoted as $e_{uv} = D(u) \cdot t(e) + D(u) \cdot r(e) + [D(u) + D(v)] \cdot f(e)$. E_v represents the energy level of node v. For $\forall u \in S_l$, $F_u \in S_{l-1}$ represents the backup nodes set for the forwarding nodes, which can directly communicate with node u, as shown by the shadow of **Fig. 2**.
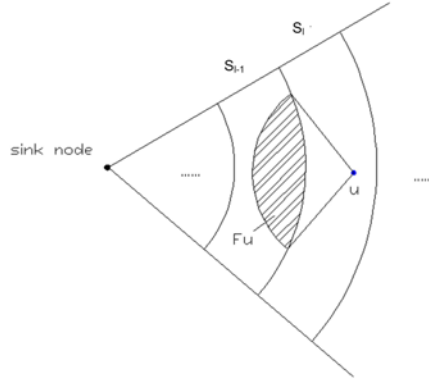
Fig. 2. Backup nodes set for the forwarding nodes

In equation (2), the first constraint specifies that the source node u in S_l has only one forwarding node. The second constraint is to guarantee that the forwarding nodes are the nodes with the highest energy level among the backup nodes.

The above optimization problem is not easily solved because of the conflicts between energy equilibrium and energy conservation. Although some of the nodes can guarantee the lowest energy consumption of the network, the remaining energy of these nodes may be low

and result in their quick death. For avoiding energy holes, we prioritize the object of energy equilibrium and this is described in detail in the following section.

4. Multi-attribute Data Fusion

4.1 Data fusion in homogeneous and heterogeneous WSNs

According to the configuration of sensor nodes, wireless sensor networks can be divided into two types: homogeneous and heterogeneous WSNs. For the former, all the sensor nodes are identical. For the latter, there are various individual differences among nodes. In our research work, the heterogeneous WSN is composed of nodes equipped with different sensors. In the following, we use an example to explain the fusion process in homogeneous and heterogeneous WSNs.

In homogeneous WSNs, the attributes, length and structure of packets generated by each sensor node are identical. As shown in **Fig. 3**, each node generates three kinds of attributive data, represented by types A, B and C. In such a case, the components with the same attributes in the packets can be fused according to the requirements set by users in data gathering.
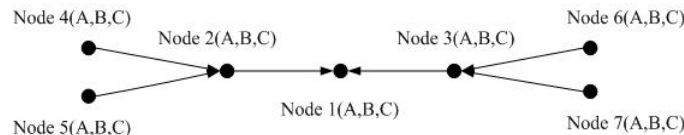

Fig. 3. Data fusion in homogeneous network

In heterogeneous WSNs, the attributes of the packets generated by different nodes are not the same, which increases the difficulty of the fusion process. As shown in **Fig. 4**, there are four kinds of sensors in the network, which can separately generate four kinds of attributive data represented by types A, B, C and D. Each node is equipped with three kinds of sensors. **Fig. 4** shows the data gathering and fusion process for nodes 5, 6, 7 and 8. The attributive data in packets generated by them are types A, B and C; types A, C and D; types B, C and D; and types   A, B and D. For easy illustration, the packets generated by nodes 1, 2, 3 and 4 are not shown in this figure. Although the data attributes of these nodes are not identical, there is still some redundancy if some of the same attributes exist. For example, node 3 can fuse the components of type A and C which coexist in packets generated by node 5 and 6. This shows that data fusion can still be used to reduce redundancy in a heterogeneous WSN if the same attributes exist in different packets.
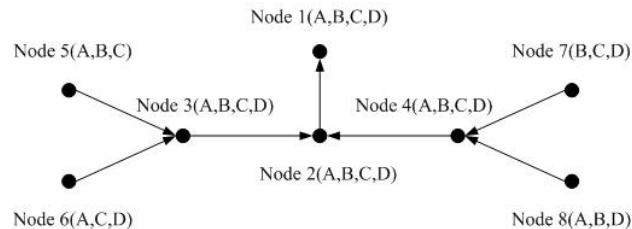

Fig. 4. Data fusion in heterogeneous network

4.2 Self-adaptive Threshold

The standard basic method used in data fusion is setting a threshold. By this means, the sensor node can decide whether to transmit the packets. If the difference between the last transmitted packet and the current prepared packet is less than the threshold, then the current packet will not be transmitted. Hence, unnecessary data transmission can be avoided.

As mentioned above, the sensor node often needs to be equipped with more than one kind of sensor, which can gather different monitoring data. Compared with separate transmission of each attributive data, the transmission of multi-attribute data packed in one packet can reduce the network flow. If all attributive data in packets still adopts the same threshold, the data with a low change rate will also be transmitted along with the packet when the threshold is low and this will waste the network energy and bandwidth. If the threshold is high, the packets will be transmitted at a low frequency and this might result in the loss of valuable data. To solve this problem, we design a self-adaptive threshold to balance the change rate of different attributive data.

We use T_1 and T_2 to represent the upper and lower limits of the self-adaptive threshold, respectively, which can be set according to the real application. In data gathering, the self-adaptive threshold of i attributive data, denoted by T_i, is calculated as:

$$T_i = T_1 + (1 - I_i) \times \frac{\sum M_i (T_2 - T_1)}{M} \tag{3}$$

Where, M is the total transmission number for the packets, M_i represents the number of packets transmitted only as a result of the i attributive data, and I_i represents the importance of the i attributive data, which can be set according to the real application in the range [0, 1]. $\sum M_i / M$ reflects the change rate of attributive i data. To calculate T_i, the sensor nodes need to cache each packet for a certain period. Obviously, the attributive data with a higher change rate will obtain a higher threshold, which leads to a stronger restriction. By this means, the passive transmission of insensitive data can be reduced. It should be noted that the arrangement of the threshold value must be restricted to avoid the loss of available data. For example, if all packet transmissions are the result of only one kind of attributive data, then $M = \sum M_i$. Here, the threshold of this attributive data is calculated by equation (4):

$$T_i = T_1 + (1 - I_i)(T_2 - T_1) = (1 - I_i)T_2 + I_i T_1 \tag{4}$$

It can be seen from equation (4) that if T_2 is set too high, and then some available data might be lost. On the contrary, if T_2 is set too low, the efficiency of the threshold to balance different attributive data will be reduced.

5. Energy-conservation Efficiency of Multi-attribute Data Fusion

In WSNs, the main task of data fusion is to reduce unnecessary data transmission for the sake of reducing energy consumption. The energy-conservation efficiency of data fusion is mainly dependent on the correlations among different data. Here, the correlation coefficient is denoted by ρ. For example, in the case of fusing two packets, the length of the generated

packet after fusion will be decreased with increasing data correlations in the two packets. Additionally, it is affected by the energy consumption of the fusion process. Especially, data fusion cannot continue to save network energy after the energy consumption reaches a certain point. In the following section, we will analyze the energy-conservation efficiency of data fusion.

To measure the energy-conservation efficiency of data fusion, both the cost of transmission and fusion are considered, denoted by $T(e_k)$ and $F(e_k)$, respectively. $T(e_k)$ is the energy consumption of a transmission on edge e_k and can be calculated as:

$$T(e_k) = G(e_k) \cdot t(e_k) \tag{5}$$

Where $G(e_k)$ represents the total data amount for transmission, and $t(e_k)$ represents the energy consumption of transmitting one-bit data. The value of $t(e_k)$ is given by the model in section 3.3.

$F(e_k)$ represents the energy consumption of the fusion process and is given by:

$$F(e_k) = H(e_k) \cdot f(e_k) \tag{6}$$

Where $H(e_k)$ represents the total data amount for data fusion, and $f(e_k)$ represents the energy consumption of fusing one-bit data. The value of $f(e_k)$ is affected by the chosen fusion algorithm.

Without loss of generality, we make the following hypothesis: When sensor node u transmits packet to node v, node v is responsible for fusing its own data and the received data. $D(u)$ and $D(v)$ represents the data amount for node u and v before fusion, respectively. $\hat{D}(v)$ represents the generated data amount after fusion. If node u and v generate one attributive data that is the same, and the correlation of node u and v is ρ_{uv}, we have:

$$\hat{D}(v) = \max(D(u), D(v)) + \min(D(u), D(v)) \cdot (1 - \rho_{uv}) \tag{7}$$

In our research work, for any $e_k \in E$, $t(e_k) = t(e)$ and $f(e_k) = f(e)$. The packet of node u sent to v includes m kinds of attributive data, so $D(u)$ can be written as:

$$D(u) = \sum_{i=1}^{m} D(u_i) \tag{8}$$

The data generated by node v includes n kinds of attributive data, so $D(v)$ can be written as:

$$D(v) = \sum_{j=1}^{n} D(v_j) \tag{9}$$

According to whether node v proceeds with data fusion or not, there are two cases. One is the case when node v does not fuse data but transmits the packets received and generated by itself directly, thus the total data amount transmitted by node v denoted as $D'(v)$ can be written as:

$$D'(v) = D(u) + D(v) = \sum_{i=1}^{m} D(u_i) + \sum_{j=1}^{n} D(v_j) \tag{10}$$

At this time, the energy consumption of node v denoted as $E_1(v)$, can be expressed by equation (11):

$$E_1(v) = D'(v) \cdot t(e) = \left(\sum_{i=1}^{m} D(u_i) + \sum_{j=1}^{n} D(v_j) \right) \cdot t(e) \tag{11}$$

On the contrary, when node v fuses the received and local packets, according to the data fusion model in section 3.2, the total data amount after fusion can be expressed by equation (12):

$$\hat{D}(v) = \sum_{i=1}^{m} D(u_i) + \sum_{j=1}^{n} D(v_j) - \sum_{i=j} \min[D(u_i), D(v_j)] \cdot \rho_{uv} \cdot g_i \tag{12}$$

Where, g_i represents the effect of the data attributes on the correlations. Obviously, $\hat{D}(v)$ decreases with increasing ρ_{uv}, which causes a reduction of energy consumption for the transmission of node v. At this time, the total energy consumption of sensor node v can be written as:

$$E_2(v) = \left[\sum_{i=1}^{m} D(u_i) + \sum_{j=1}^{n} D(v_j) \right] \cdot f(e) + \left[\sum_{i=1}^{m} D(u_i) + \sum_{j=1}^{n} D(v_j) - \sum_{i=j} \min[D(u_i), D(v_j)] \cdot \rho_{uv} \cdot g_i \right] \cdot t(e)$$

$$= \left[\sum_{i=1}^{m} D(u_i) + \sum_{j=1}^{n} D(v_j) \right] \cdot [t(e) + f(e)] - \sum_{i=j} \min[D(u_i), D(v_j)] \cdot \rho_{uv} \cdot g_i \cdot t(e) \tag{13}$$

It can be deduced that the energy-conservation efficiency of multi-attribute fusion at node v is:

$$U(v) = E_1(v) - E_2(v)$$

$$= \sum_{i=j} \min[D(u_i), D(v_j)] \cdot \rho_{uv} \cdot g_i \cdot t(e) - \left(\sum_{i=1}^{m} D(u_i) + \sum_{j=1}^{n} D(v_j) \right) \cdot f(e) \tag{14}$$

Above all, the energy-conservation efficiency of data fusion is determined by both the correlations among nodes and the energy consumption of the fusion process. If the energy consumption of fusion is too high or the correlations among nodes are too low, data fusion may not save energy for WSNs. The energy-conservation efficiency of data fusion can be used to optimize the establishment of routing and determine how to proceed with data fusion.

It should be noticed that we are actually measuring the energy consumption between two nodes capable of direct communication. In fact, there are unavoidable packet collisions and losses, which will cause data retransmission and thus increase energy consumption. In our research work, we are less concerned with the quality of communication, so we ignore the effects of packet collisions and packet losses. An increase in energy consumption caused by packet retransmission can be equivalently measured by increasing $t(e)$. Additionally, the cost of computing the energy-conservation efficiency is also ignored, since this cost is much lower compared to the complicated fusion process and will not influence the overall performance of WSNs.

6. Multi-attribute Fusion Tree

In this section, we proposed an energy equilibrium routing method for WSN, viz., multi-attribute fusion tree (MAFT). The establishment of MAFT is determined by the remaining energy of sensor nodes and the energy-conservation efficiency of data fusion. The purpose of our design is to balance and save energy in WSNs.

6.1 Multi-attribute fusion tree

In wireless sensor networks, the routing method is responsible for gathering the information from the sensor nodes to the sink node. Each sensor node can play a dual functional role such as relays and terminals in traditional networks. Since both the energy supply and computing capacity are limited, the main target for routing research is focusing on how to establish a high efficiency routing method to maximize the network lifetime. Additionally, the establishment of routing should have the characteristics of low complexity and serviceability.
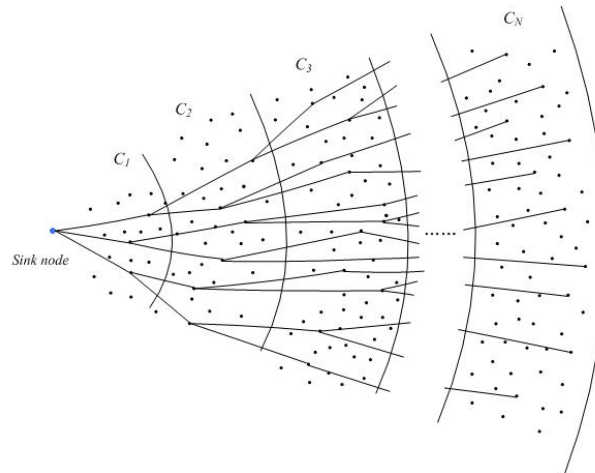
Fig. 5. Division of sensor nodes hops in network

For these reasons, we establish a multi-attribute fusion tree (MAFT). The network is divided into N coronas, denoted by $C_1, C_2, C_3, \cdots, C_N$ as shown in **Fig. 5**. For the sake of eliminating packet losses, the width of the coronas should be less than the maximum communication range of the nodes. In data gathering, nodes belonging to a corona $\{C_{i-1} | i \neq R\}$ will forward packets generated by both themselves and nodes from corona $\{C_j | (i+1) \leq j \leq R\}$.

Each sensor node in corona C_i ($2 \leq i \leq N$) can find its forwarding node from the corona C_{i-1} while the nodes in corona C_1 can communicate with the sink node directly. Each forwarding node needs to cache the received packets for a certain period. This routing method can be constructed only with local information by a distributed process. The location of each node is not important in this process. Additionally, MAFT can support both active and passive data gathering. In active data gathering, each sensor node transmits packets by pre-installed rules. Unlike active data gathering, each sensor node in passive data gathering will not transmit the packets until it receives an inquire message.

6.2 Problem Optimization

If energy equilibrium is not considered in equation (2), it means the second constraint is ignored. The object of optimization is simply to achieve maximal network energy savings. According to the simplified object of optimization, the remaining energy of the selected forwarding nodes may be low and more tasks may need to undertaken. Although the total energy consumption of the whole network is lowest during a certain time, partial nodes that consume energy first will speed up the death of the network. So, the energy equilibrium must first be considered.

However, the absolute energy equilibrium cannot be realized in a real application. We adopt a simple method to ensure some degree of energy equilibrium, which involves selecting those nodes with higher remaining energy as forwarding nodes to undertake more tasks. As described in the second constraint of equation (2), when the forwarding node of node u is to be selected, the nodes with higher remaining energy in F_u are first selected, denoted as F_u', where $F_u' \in F_u$. Then, the nodes in F_u' which can meet the needs of optimization are selected as forwarding nodes. The detailed selection process is described in section 6.3. Now, equation (2) is simplified as follows:

$$
\begin{cases}
\min \sum_{u \in S_l} \sum_{v \in F_u'} e_{uv} \cdot x_{uv}, & l = (1,2,3,\cdots m) \\
s.t \ \ for \ \forall u \in S_l, \ \ \sum_{v \in F_u'} x_{uv} = 1
\end{cases}
\tag{15}
$$

Theorem 1: The bigger $U(v)$ becomes, the smaller $\sum_{u \in S_l} \sum_{v \in F_u'} e_{uv} \cdot x_{uv}$ becomes.

Proof: $\sum_{u \in S_l} \sum_{v \in F_u} e_{uv} \cdot x_{uv}$ represents the total energy consumption of sending the data generated by all sensor nodes to the sink node. If there is no fusion process, the total energy consumption will be the highest. If the energy consumption of the sink node is ignored, the energy consumption for sending the data generated by $\forall u \in S_l$ to the sink node can be calculated as:

$$
e_u = l \cdot D(u) \cdot t(e) + (l-1) \cdot D(u) \cdot r(e)
\tag{16}
$$

The $\sum_{u \in S_l} \sum_{v \in F_u} e_{uv} \cdot x_{uv} = \sum_{u \in S_l} e_u = \sum_{u \in S_l} [l \cdot D(u) \cdot t(e) + (l-1) \cdot D(u) \cdot r(e)]$
\qquad (17)

Obviously, since the hop number is constant for each node, $\sum\limits_{u \in S_l}\sum\limits_{v \in F_u} e_{uv} \cdot x_{uv}$ is determined by the total data amount. After employing the data fusion process, the data of node u will be fused at the forwarding node v. As the correlations of nodes in one-hop range are the biggest, the total data amount decreases with decreasing $\hat{D}(u)$, so the energy consumption also decreases.

From equation (11) and (13), it can be deduced that:

$$U(v) = [D(u) + D(v)] \cdot [t(e) - f(e)] - \hat{D}(v) \cdot t(e) \tag{18}$$

Obviously, the bigger $U(v)$ becomes, the smaller $\hat{D}(v)$ becomes, thus, the smaller $\sum\limits_{u \in S_l}\sum\limits_{v \in F_u} e_{uv} \cdot x_{uv}$ becomes. Theorem 1 is proven.

6.3 Selection of Next-hop Forwarding Nodes

As illustrated above, it is suggested that the selected forwarding nodes from the backup nodes with higher remaining energy ought to maximize energy-conservation efficiency, since each sensor node might have more than one backup node chosen as a forwarding node. Assume that there are H backup nodes for node u, denoted by $\{v_1, v_2, v_3, \cdots, v_H\}$. Now, the most important thing is to select the proper forwarding nodes from these. As described in section 6.2, both the remaining energy and the energy-conservation efficiency are considered during the selection. We define the energy-conservation efficiency of multi-attribute fusion as $U(v_j)$ and it can be calculated by each sensor node using equation (13). If we only consider this, the backup node with the highest $U(v_j)$ will be chosen. For example, if one forwarding node represented by v_r is selected, it should satisfy equation (19):

$$U(v_r) \geq U(v_j) \qquad v_r \in H \tag{19}$$

However, this will violate energy equilibrium in WSNs. Because sensor nodes in WSNs are usually stable, the data correlations of nodes are also fixed so that the $U(v_j)$ of each backup node seldom changes. This means that v_j will be chosen as the forwarding node repeatedly, which increases its energy consumption and accelerates its death. To avoid energy holes appearing in WSNs, the remaining energy of the nodes must be considered first.

As the remaining energy of nodes decreases continuously, the information cannot be exchanged frequently. We divide the initial energy E_0 into grades to measure the remaining energy of nodes. Using the energy grade, the sensor node needs to inform its neighbor nodes when its energy grade decreases, which means each node needs to update its energy grade information no more than $G-1$ times. Obviously, this method can significantly reduce the energy consumption of communication.

When all backup nodes are determined, it first compares the energy grades of the backup nodes, and then selects the node with the highest one. If there is more than one backup node with the highest energy grade, further selection is based on $U(v_j)$. If there are still multiple

nodes with the highest $U(v_j)$ after further selection, then a random node among them will be selected.

6.4 Establishment of MAFT

The establishment of MAFT is started from the sink node by broadcasting an advertisement (ADV) message. The nodes receiving this message belong to corona C_1. Then the nodes in corona C_1 broadcast their ADVs including their ID information, corona serial number and current energy grade. The nodes receiving the message from corona C_1 belong to corona C_2. The process is continued until the whole network is divided into N concentric coronas. In broadcasting transmission, nodes in corona C_i are unavoidable to receive ADV messages from nodes in the same or outer corona C_{i+1}. To avoid duplicate recording, the sensor nodes need to compare their corona serial number with those in ADV, and abandon the ADV from corona C_i and C_{i+1}.

For further energy balancing, we divide each corona into many sub-zones, as shown in **Fig. 6**. The data is transmitted between the corresponding sub-zones in the two neighbor coronas. For example, the backup nodes set F_u for node u are all the nodes in the pink area.
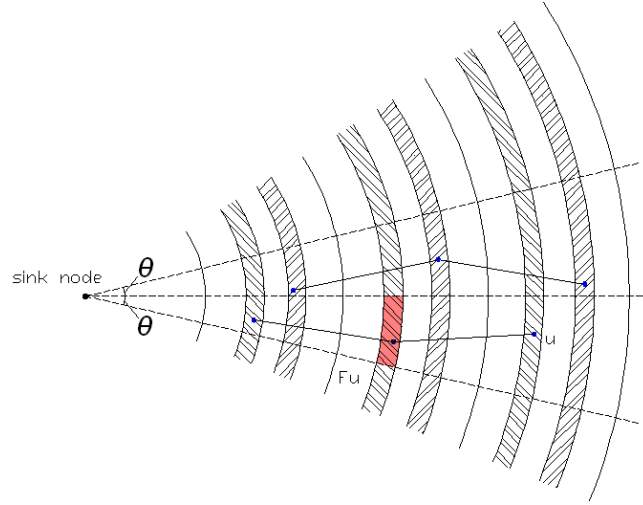

Fig. 6. MAFT

Generally, the sensor nodes will receive more than one ADV message from the inner corona. These nodes will become the backup nodes for the forwarding nodes. The selection process of the forwarding nodes starts from the nodes in corona C_N to the sink node. The selection results should be informed to the chosen nodes.

Fig. 7 shows the selection process of the forwarding nodes. $N_f(u)$ represents the forwarding node of node u. $M(v)$ represents the ADV message of node v, including $(ID_v, S_v, G(v))$.

In WSNs, it is impossible for the energy consumption to be in absolute equilibrium. Therefore, the routing needs to be reconstructed after a certain period, which depends on many factors, such as the application environment, the initial energy level of sensor nodes, etc.

```
1: Initialize i=0
2: for ∀u∈S do
3:    if M(v)=True, Su=Sv+1 then
4:        Fu←v
5:      for all v∈Fu do
6:        if G(v)>G(z), ∀z∈Fu then
7:            i++
8:          denote vi= v
9:        end if
10:     end for
11:    if i=1 then
12:        Nf(u)=vi
13:    else
14:        calculate U(vi), j=1,2,…,i
15:        Nf(u)=argmaxU(vi)
16:    end if
17: end
```

Fig. 7. Selection process of forwarding nodes

7. Performance Evaluation

In this section, we evaluate the performance of MAFT via simulation experiments under NS2. We assume that 300 sensor nodes are uniformly deployed in a circular area with a diameter of 200m. All the sensor nodes are information sources and can be forwarding nodes. The original packet size is 500 bytes and is generated by each sensor node. Each packet consists of three kinds of attributive data. The basic parameters used in the simulations are shown in **Table 1**.

Table 1. Simulation parameters

| Parameter	Value
Initial energy	2J
Communication bandwidth	1Mbps
Delay time	25μs
Energy consumption/circuit	50nJ/bit
Energy consumption of amplifier	$d<87$m, 10 pJ/bit·m^2 $d\geq87$m, 0.0013 pJ/bit·m^4
d_c	20m
d_s	25m
Energy consumption of data fusion	20nJ/bit

We employ d_s and d to represent the maximum correlation distance between nodes and the space distance between two sensor nodes, respectively. If $d \geq d_s$, the correlation coefficient between nodes is 0, otherwise, it is given by $\rho = (1 - d/d_s) \times f \times g_i$. In our experiments, the corresponding values of the three given data attributes in the packets are $g_i = 0.6, 0.8, 1.0 (i = 1,2,3)$. f represents the effect of the fusion algorithm on the data correlations, which is set as 1 in the whole simulation. We compared MAFT with MTE-F, where MTE-F includes the data fusion process based on MTE. The main idea of minimum transmitted energy (MTE) is that all nodes send packets using the minimum power, that is to say, they send packets to their nearest neighbor nodes. For further analysis, we also include

MAFT-NE and MAFT-NF for comparison. Compared with MAFT, the main difference is that the remaining energy is not considered in MAFT-NE, while the energy-conservation efficiency of data fusion is not considered in MAFT-NF.
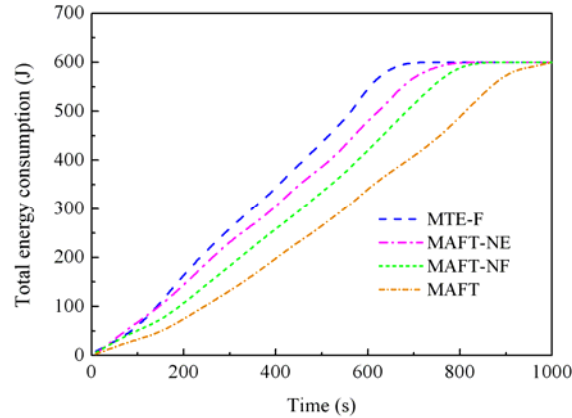
**Fig. 8**. Energy consumption of network.

Fig. 8 shows the total energy consumption under different protocols. It can be seen that the total energy consumption in MAFT is lowest when both the energy-conservation efficiency of data fusion and the remaining energy is considered. MAFT-NF shows more total energy consumption than MAFT, which does not consider the energy-conservation efficiency of data fusion. Compared to MAFT and MAFT-NF, MAFT-NE shows much higher total energy consumption when the remaining energy is not considered. MTE-F shows the highest total energy consumption among the four protocols.
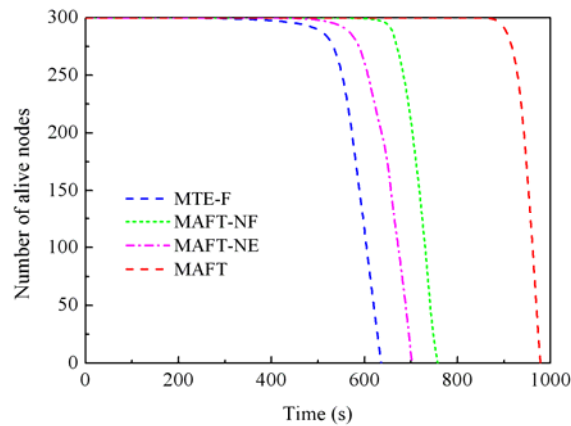

Fig. 9. Number of live nodes

Fig. 9 shows the number of live nodes with different protocols. In MAFT, the dead times of the first and last node are both later than the others. MAFT-NF is sensitive to the energy changes of the network. Hence, it shows better performance than MAFT-NE. MTE-F still shows the worst performance.
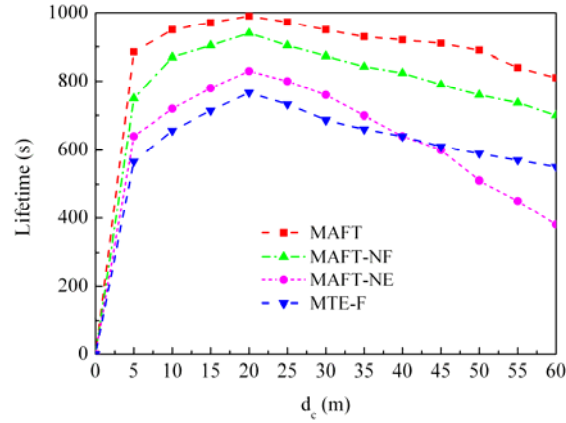
Fig. 10. Effect of d_c on network lifetime.

Fig. 10 reflects the effect of d_c on the network lifetime with different protocols, where d_s is fixed at 25m. It can be seen that the lifetime of MAFT is always longest at the same d_c and the peak appears when d_c is 20m. When d_c is less than 20m, the number of routing hops between the sensor nodes and the sink node increases with decreasing d_c. When d_c is over 20m and less than 25m, the correlation coefficient ρ decreases with increasing d_c. When d_c has reached 25m, the correlation coefficients ρ of MAFT, MAFT-NE and MAFT-NF have reached zero. As MAFT-NE only considers the fusion efficiency, the lifetime of MAFT-NE shows the most rapid decline. Since data packages are transmitted to neighbor nodes in MTE-F, the real communication distance is not determined by d_c. The fusion process is still available in MTE-F, which results in a slow decline in lifetime. It can be seen that the lifetime of MTE-F is higher than that of MAFT-NE when d_c is higher than 40m.
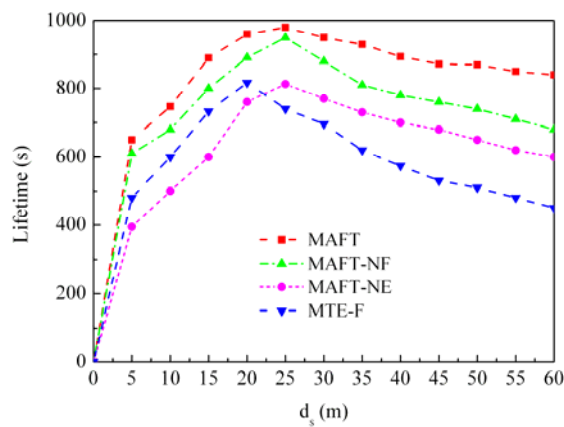
Fig. 11. Effect of d_s on network lifetime

Fig. 11 reflects the effect of d_s on the network lifetime with different protocols, where d_c is fixed at 20m. When d_s is less than 20m, the energy consumption will not benefit from data fusion in data gathering, since ρ is now 0. However, MAFT and MAFT-NF still exhibits a longer lifetime than the other two protocols. MTE-F shows better performance than MAFT-NE, because of the higher sensitivity of MAFT-NE over MTE-F to the effect of d_s on the correlation coefficients. On the contrary, when d_s is higher than 20m, ρ decreases with increasing d_s. At this time, MAFT still exhibits the best performance.

8. Conclusions

As there are strong correlations between data gathered from sensor nodes in close physical proximity, effective in-network fusion schemes involve minimizing such redundancy and hence reducing the load in wireless sensor networks. Therefore, the routing method should support data fusion for fusion-driven routing. Here, data is fused along the routing path toward the sink node. Under a complicated environment, the data gathered must be multi-attribute for each sensor node equipped with more than one kind of sensor. An increase in the complexity for the fusion process is unavoidable due to the existence of various physical attributes.

In this paper, we analyze the process and performance of multi-attribute fusion in data gathering for homogeneous and heterogeneous WSNs. Based on the analysis we propose a self-adaptive threshold to balance the different change rates of multi-attribute data in the fusion process. Next, the energy-conservation efficiency of multi-attribute fusion in data gathering is discussed. We present a method to measure the energy-conservation efficiency of multi-attribute fusion that considers both the effect of multi-attribute and the cost of data fusion. Then we propose an energy equilibrium routing method, viz., multi-attribute fusion tree (MAFT). In MAFT, both energy equilibrium and conservation are considered. Finally, we perform extensive simulation experiments to evaluate MAFT by several performance criteria. The results show that MAFT can save energy and prolong the network lifetime.

Acknowledgement

References

[1]   T. Shon, B. Koo, H. Choi, Y. Park, "Security architecture for IEEE 802.15.4-based wireless sensor network," in *4th International Symposium on Wireless and Pervasive Computing*, Melbourne, VIC, Australia, Inst. of Elec. and Elec. Eng. Computer Society, 2009.

[2]   P. K. Sahoo, K. Hsieh, J. Sheu, "Boundary node selection and target detection in wireless sensor network," in *4th IEEE and IFIP International Conference on Wireless and Optical Communications Networks*, Singapore, Singapore, Inst. of Elec. and Elec. Eng. Computer Society, 2007.

[3]   N. A. Pantazis, D. D. Vergados, N. I. Miridakis, D. J. Vergados, "Power Control Schemes in Wireless Sensor Networks for homecare e-Health Applications," in *1st International Conference*

on Pervasive Technologies Related to Assistive Environments, Athens, Greece, Assoc. Comput. Mach., pp. 1100-1107, 2008.

[4]   J. Li, P. Mohapatra, "Analytical modeling and mitigation techniques for the energy hole problem in sensor networks", *Pervasive and Mobile Computing*, vol. 3, pp. 233-254, 2007.

[5]   A. P. Ruth, M. Jason, B. Abu, K. Ashok, B. Magdy, "Multisensor data fusion schemes for wireless sensor networks", in *CAMPS 2006-International Workshop on Computer Architecture for Machine Perception and Sensing*, Montreal, Canada, Inst. of Elec. and Eng. Computer Society, pp.136-141, 2006.

[6]   J. Wang, H. Yu, Z. Shang, "Research on Reliable Link Layer Communication in Wireless Sensor Networks," in *Proc. of the 2005 International Conference on Communications, Circuits and Systems. Piscataway: IEEE Press*, pp. 417-421, 2005.

[7]   Y. Wei, J. Heidemann, D. Estrin, "Medium access control with coordinated adaptive sleeping for wireless sensor networks", *IEEE/ACM Transactions on networking*, vol.12, no. 3, pp.493-506, 2004.

[8]   B. Krishnamachari, "The Impact of Data Aggregation in Wireless Sensor Networks," in *Proc. of International Workshop on Distributed Event-Based Systems*. pp. 575-578, 2002.

[9]   C. Intanagonwiwat, D. Estrin, R. Govindan, J. Heidemann, "Impact of network density on data aggregation in wireless sensor networks," in *22nd International Conference on Distributed Systems,* Vienna, Austria: Institute of Electrical and Electronics Engineers Inc., pp.457-458, 2002.

[10] S. Pattem, B. Krishnamachari, R. Govindan, "The impact of spatial correlation on routing with compression in wireless sensor networks," in *Third International Symposium on Information Processing in Sensor Networks,* Berkeley, CA., United states: Association for Computing Machinery, pp.28-35, 2004.

[11] Y. Yu, B. Krishnamachari, V. K. Prasanna, "Energy-latency tradeoffs for data gathering in wireless sensor networks," in *IEEE INFOCOM 2004 - Conference on Computer Communications - Twenty-Third Annual Joint Conference of the IEEE Computer and Communications Societies*, Hongkong, China, pp.244-255, 2004.

[12] A. Goel, D. Estrin, "Simultaneous optimization for concave costs: Single sink aggregation or single source buy-at-bulk," in *Proc. of ACM-SIAM Symposium on Discrete Algorithms*, Baltimore, MD, 2003.

[13] R. Cristescu, B. Beferull-Lozano, M. Vetterli, "On network correlated data gathering," *In Proceedings of IEEE Infocom*, Hongkong, China, 2004.

[14] H. Luo, Y. Liu, S. K. Das, "Routing Correlated Data with Fusion Cost in Wireless Sensor Networks," *IEEE Transactions on Mobile Computing*, vol. 5, no. 11, pp.1620-1632, 2006.

[15] H. Luo, J. Luo, Y. Liu, S. K. Das, "Adaptive data fusion for energy efficient routing in wireless sensor networks," *IEEE Transactions on Computers*, vol. 55, no. 10, pp.1286-1299, 2006.

[16] P.V. Rickenbach, R. Wattenhofer, "Gathering correlated data in sensor networks," in *Proc. ACM Joint Workshop Foundations of Mobile Computing (DIALM-POMC '04)*, 2004.

[17] A. Goel, K. Munagala, "Balancing steiner trees and shortest path trees online," in *Proc. 11th Ann. ACM-SIAM Symp, Discrete Algorithms (SODA '00)*, 2000.

[18] H. Luo, Y. Liu, S. K. Das, "Distributed algorithm for en route aggregation decision in wireless sensor networks," *IEEE Transactions on Mobile Computing,* vol. 8, no. 1, pp.1-13, 2009.

[19] A. Anandkumar, W. Meng, T. Lang, A. Swami, "Prize-Collecting Data for Cost-Performance Tradeoff Distributed Inference," in *Proc. IEEE INFOCOM*, pp.2150-2158, 2009.

[20] G. Xing, R. Tan, B. Liu, Wang J., Jia X., Yi C., "Data Fusion Improves the Coverage of Wireless Sensor Networks," in *Proc. of International Conference on Mobile Computing and Networking*, pp.157-168, 2009.

[21] W. B. Heinzelman, A. P. Chandrakasan, H. Balakrishnan, "An Application-Specific Protocol Architecture for Wireless Microsensor Networks," *IEEE transaction on wireless communications*, pp.660-670, 2002.

Kai Lin is a lecturer at the School of Electronic and Information Engineering, Dalian University of Technology. He received his B.S. degree in the School of Electronic and Information Engineering, Dalian University of Technology, Dalian, China in 2001, and obtained his M.S. and PhD degrees from the College of Information Science and Engineering, Northeastern University, Shenyang, China in 2005 and 2008, respectively. His research interests include wireless networks, ubiquitous computing and embedded technology.

Lei Wang is an associate professor at the School of Software of Dalian University Technology. He received his B.S., M.S., and PhD degrees in Tianjin University in 1995, 1998 and 2001, respectively. He was a Senior Research Fellow of Bell Lab in Alcatel Lucent, Beijing, from 2001 to 2004. Then he worked for two years in the Digital Technology Research Institute of Samsung. He worked as a postdoc in Washington State University in 2007, USA. His research interests include wireless sensor networks and wireless ad hoc networks.

Keqiu Li received his Bachelor and Master degrees in the Department of Applied Mathematics, Dalian University of Technology, Dalian, China in 1994 and 1997, respectively. He obtained his PhD degree from the Graduate School of Information Science, Japan Advanced Institute of Science and Technology, Japan. He is currently a professor at the School of Electronic and Information Engineering, Dalian University of Technology, China. His research interests include wireless sensor networks, computing and peer-to-peer computing.

Lei Shu is a currently a Specially Assigned Research Fellow in the Department of Multimedia Engineering, Graduate School of Information Science and Technology, Osaka University, Japan. He received his B.S. degree in Computer Science from South Central University for Nationalities, China, 2002, his M.S. degree in Computer Engineering from Kyung Hee University, Korea, 2005, and his PhD degree in the Digital Enterprise Research Institute, NUIG, in 2010. He is a member of the Libelium WSNs group, ACM and IEEE. (IEEE membership ID: 90411188). Also, he has served as editor of 19 international journals. To date, he has published over 80 papers in related conferences, journals and books. His research interests include semantic sensor networks, wireless sensor networks, context aware and sensor network middleware.