IEICE TRANS. COMMUN., VOL.E88–B, NO.12 DECEMBER 2005

PAPER *IEICE/IEEE Joint Special Section on Autonomous Decentralized Systems*

# Energy-Efficient Data-Centric Routing in Wireless Sensor Networks*

**Hong-Hsu YEN**[†a)], ***Member***, **Frank Yeong-Sung LIN**[††b)], ***and*** **Shu-Ping LIN**[††c)], ***Nonmembers***

**SUMMARY** Incorporating sensor nodes with data aggregation capability to transmit less data flow in wireless sensor networks could reduce the total energy consumption. This calls for the efficient and effective data-centric routing algorithm to facilitate this advantage. In the first part of this paper, we model the data-centric routing problem by rigorous mixed integer and linear mathematical formulation, where the objective function is to minimize the total transmission cost subject to multicast tree constraints. With the advancement of sensor network technology, sensor nodes with configurable transmission radius capability could further reduce energy consumption. The second part of this paper considers the transmission radius assignment of each sensor node and the data-centric routing assignment jointly. The objective function is to minimize the total power consumption together with consideration of construction of a data aggregation tree and sensor node transmission radius assignment. The solution approach is based on Lagrangean relaxation in conjunction with the novel optimization-based heuristics. From the computational experiments, it is shown that the proposed algorithms calculate better solution than other existing heuristics with improvement ratio up to 169% and 59% with respect to fixed transmission radius and configurable transmission radius for network with 300 random generated nodes.

***key words:*** *energy saving, data aggregation, data-centric routing, Lagrangean relaxation, wireless sensor networks*

## 1. Introduction

The wireless sensor networks are types of nascent technologies that probe and collect environmental information, such as temperature, atmospheric pressure and irradiation by providing ubiquitous sensing, computing and wireless communication capabilities. Wireless sensor networks are similar to mobile ad-hoc networks (MANETs) in that both involve multi-hop communications. However there are two main differences. First, typical communication mode in wireless sensor networks is to transmit data from multiple data source nodes to one data sink node. This is a kind of *reverse-multicast* rather than communication between any pair of nodes in MANETs. Second, since data are collected by multiple sensors there must be some redundancy in reported data, which are being transmitted by numerous sources. This would rapidly deplete the energy of sensors and result in disconnected network. Data aggregation, therefore, has been put forward as a particularly useful function for routing in terms of energy consumption in wireless sensor networks [5], [6].

Sensor nodes are usually scattered in a *sensor field*. When any event occurs, such like surging irradiation or temperature declining below certain threshold, sensor nodes within specific *sensing range* detect this event and collect the data which would be transmitted to the sink node for taking further processing. We refer to each sensor node as *data source* since data are generated from sensors, and the sink node as *data sink*. The application scenario described above is called *event-driven* model in which sensors are assigned to detect a particular event. There are two other different applications of wireless sensor networks, namely, *periodic* and *query-based*. In *periodic* scenario, sensors probe environmental information periodically and report their measurements back to the sink node. All sensors in this kind of networks are necessitated to be synchronized such that all sensors sense information and report it simultaneously. Query-based scenario is applied to user-oriented applications. User can query information from certain area of sensors to acquire measurements that user interested in.

In event-driven model if specific event happens, raw data are collected and processed before transmission. Redundant and useless data are discarded. The local raw data are first combined together and the aggregated result is transmitted to sink node. Interestingly, data are routed along *reverse multicast tree* where multiple data sources transmit information back to the sink node (e.g. Fig. 2). Every non-leaf node on this reverse multicast tree could perform data aggregation function to summarize the outputs from downstream data sources. This process is called *data-centric routing*.

Data aggregation is the key to the data-centric routing, not only combining the data coming from different sources and eliminating redundancy, but also minimizing the total number of transmissions involved in data routing in such a way to save energy of sensors. In addition to redundancy suppression, other aggregation function could be MAX, MIN, and SUM. In this paper we assume that every node posses data aggregation capability, which transmits a single aggregated packet if it receives multiple input packets to the same data sink. Figure 1 gives an example of data-centric routing where the average temperature is reported to the data sink. The aggregation function is AVG. Label $x(y)$ at each node represents the local temperature measurement

Manuscript received March 30, 2005.

†The author is with the Department of Information Management, Shih Hsin University, Taiwan.

††The authors are with the Department of Information Management, National Taiwan University, Taiwan.

*Earlier version of this paper appears in Proc. of IEEE ICC 2005.

a) E-mail: hhyen@cc.shu.edu.tw

b) E-mail: yslin@im.ntu.edu.tw

c) E-mail: r92018@im.ntu.edu.tw

DOI: 10.1093/ietcom/e88–b.12.4470

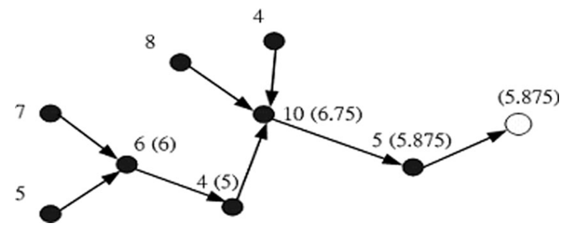
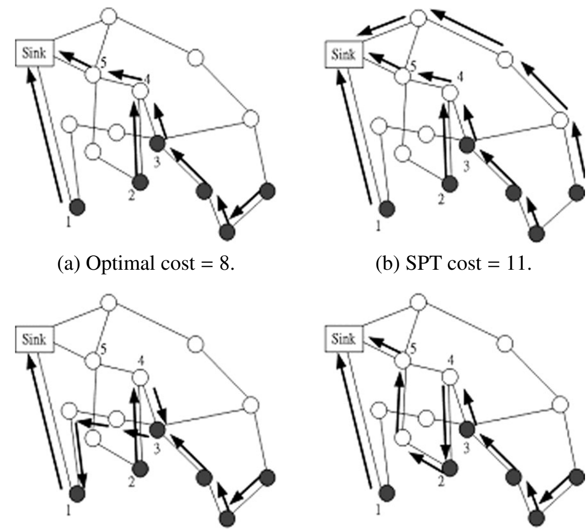



**Fig. 1** Illustrative example of data aggregation.




(a) Optimal cost = 8. (b) SPT cost = 11.

(c) CNS cost = 9. (d) GIT cost = 9.

**Fig. 2** Simple illustrative example of SPT, CNS, and GIT.

is $x$ while the aggregated (average) value so far is $y$. For example, at node 4(5), the average temperature is $(4+6)/2 = 5$.

In wireless sensor networks, since the transmission power is associated with the physical distance between the source and the destination, it is reasonable to assume that the transmission cost associated with each link is identical to the transmission cost with its opposite direction. By this assumption, the total transmission cost of Fig. 1 is identical to the multicast tree transmission cost where the root is node (5.875) and the other nodes are the destinations. However, constructing the minimum cost multicast tree is the well-known Steiner tree problem, which is proven to be the NP-completeness [4]. This calls for the effective and efficient heuristic to solve this problem.

Fixed transmission radius data-centric routing problem in wireless sensor networks has been studied in existing research. S. Singh [8] shows that by using new power-aware metrics, for example energy consumed for transmitting per packet, for determining routes in wireless ad-hoc networks, shortest cost routing algorithm based on these new power-aware metrics could reduce cost/packet of routing packets over shortest hop routing. This inspires us to construct power-aware metrics ($a_l$ in Sect. 2), instead of hops which is used in [6], as the link cost. H. O. Tan [9] proposes centralized heuristic based on Prim's shortest path algorithm to construct a data aggregation tree. This heuristic incorporates residual energy of sensor nodes into Prim's algorithm in order to prolong lifetime of sensor nodes. However, as shown in our computational results, the data aggregation tree constructed by shortest path algorithm (shortest path tree, SPT) is not good enough in terms of transmission cost.

Krishnamachari [6] devises three interesting suboptimal aggregation heuristics, namely, Shortest Paths (SPT), Center at Nearest Source (CNS), and Greedy Incremental Tree (GIT), respectively. Figure 2 is a simple illustration of these three heuristics. Note that the transmission cost on each link each all set to be 1. From Fig. 2, we see that none of these three heuristics locate optimal solution. In SPT scheme, each data source node finds the shortest path back to sink node. Figure 2(b) shows the tree generated by SPT scheme. CNS selects one node that is nearest to the sink node as the aggregation node and other data source nodes connect to this aggregation node by using the shortest hop path. Figure 2(c) shows the final routing assignment by adopting CNS heuristic.

In GIT scheme, initially the member in the tree is only the sink node. Each data source finds the shortest hop path to this tree and the data sources with the minimum hop along with the intermediate nodes on this path are included in this tree. This process is repeated until all data source nodes are included in the tree. Note that how to properly select the path when there are two paths with the same hop distance to the tree will have significant impact on the solution quality of the GIT. In Fig. 2(d), after the nearest node 1 connecting to sink node, node 2 and 3 are three hops away from the tree consisting of sink node and node 1. If node 2 selects path through node 4 and 5 to reach sink node then the resultant tree will be optimal case.

The basic idea of fixed transmission radius data-centric algorithms is to save energy by reducing number of sensor nodes involved in data aggregation tree. If the transmission radius of sensor node could be configured, it is believed that energy consumption could be further reduced. The power consumption of transmitting data distance $r$ is measured as $r^{\alpha} + c$, where $\alpha$ is a signal attenuation constant (usually between 2 to 4) and $c$ is a positive constant that represents signal processing. [3] studies the tradeoff between power consumption transmission radius and coverage of transmission node. For long transmission radius, more sensor nodes could be covered such that the total number of transmission could be reduced. However, with large signal attenuation constant (e.g. 4), long transmission radius incur significant power consumption that would sacrifice the gain from reduced total number of transmission.

In this paper, we first propose an optimization-based heuristics to solve the *fixed transmission radius data-centric routing problems (DCR)* in wireless sensor networks. The problem is first formulated as a mixed integer and linear programming (MILP) problem where the objective function is to minimize the total transmission cost used for all multicast groups subject to multicast tree and data aggregation constraints. In the second part of this paper, besides routing assignment, we also study the transmission radius assignment of sensor nodes to further reduce total energy consumption. Hence, the *energy-efficient data-centric routing problem (EDCR)* in wireless sensor network could be formally defined as minimizing total power consumption subject to reverse-multicast tree, and configurable transmission radius. We propose the Lagrangean relaxation scheme in conjunction with the optimization-based heuristics to solve these two problems. From the computational experiments, the proposed solution approaches are superior to the existing heuristics.

The remainder of this paper is organized as follows. In Sect. 2, a MILP formulation of basic model of the DCR problem is proposed. In Sect. 3, solution approaches to basic model based on Lagrangean relaxation are presented. In Sect. 4 heuristics to basic model are developed for calculating good primal feasible solution. In Sect. 5, computational results of DCR problem are reported. In Sect. 6, mathematical formulation of EDCR problem is proposed. In Sect. 7, solution approaches to extension model based on Lagrangean relaxation are presented. In Sect. 8, heuristics are developed to get good primal feasible solution of EDCR problem. In Sect. 9, computational experiments of EDCR problem are reported. Finally, Sect. 10 concludes this paper.

## 2. Problem Formulation to DCR Model

A data-centric wireless sensor network is modeled as a graph in which sensors are represented as nodes and the arc connected two nodes indicates that one node is within the other's transmission radius. The definition of notations adopted in the formulation is listed below.

| | |
|---|---|
| $G$ | The set of all multicast groups |
| $D_g$ | The set of data source nodes for the multicast group $g$ |
| $L$ | The set of all links in the graph |
| $P_{gd}$ | The set of candidate paths from the data source node $d$ to the sink node of multicast group $g$ |

| | |
|---|---|
| $h_g$ | Longest distance of shortest path to reach farthest data source node for multicast group $g$ |
| $a_l$ | Unit power aware transmission cost associated with the link $l$ |
| $\delta_{pl}$ | The indicator function which is 1 if the link $l$ is on the path $p$ and 0 otherwise |

In this formulation, we generalize the formulation to consider multiple multicast groups, i.e. multiple data sink node. The decision variables for the wireless sensor networks routing problem are denoted as follows.

| | |
|---|---|
| $C_l$ | Number of data units transmitted through link $l$ |
| $y_{gl}$ | 1 if the multicast group $g$ uses the link $l$ and 0 otherwise |
| $x_{gpd}$ | 1 if the multicast group $g$ uses the path $p$ to reach the source node $d$ and 0 otherwise |

The data-centric routing problem in wireless sensor networks is then formulated as the following combinatorial optimization problem (IP).

$$Z_{IP} = \min \sum_{l \in L} a_l C_l \tag{IP}$$

subject to:

$$\sum_{g \in G} y_{gl} \le C_l \qquad \forall l \in L \tag{1}$$

$$C_l \in \{0, 1, 2, 3, ...., |G|\} \qquad \forall l \in L \tag{2}$$

$$\sum_{p \in P_{gd}} x_{gpd}\delta_{pl} \le y_{gl} \qquad \forall g \in G, l \in L, d \in D_g \tag{3}$$

$$y_{gl} = 0 \text{ or } 1 \qquad \forall g \in G, l \in L \tag{4}$$

$$\sum_{l \in L} y_{gl} \ge \max\left\{h_g, \left|D_g\right|\right\} \qquad \forall g \in G \tag{5}$$

$$\sum_{d \in D_g} \sum_{p \in P_{gd}} x_{gpd}\delta_{pl} \le \left|D_g\right| y_{gl} \qquad \forall g \in G, l \in L \tag{6}$$

$$\sum_{p \in P_{gd}} x_{gpd} = 1 \qquad \forall g \in G, d \in D_g \tag{7}$$

$$x_{gpd} = 0 \text{ or } 1 \qquad \forall p \in P_{gd}, g \in G, d \in D_g. \tag{8}$$

The objective function of (IP) is to minimize the total data transmission cost for the wireless sensor networks, which equals to the total multicast routing cost. Constraint (1) requires that the number of multicast groups adopting link $l$ on its multicast tree should be less then or equal to the number of data units transmitted through link $l$. Constraint (2) requires that number of data units on link $l$ be at most cardinality of $G$, i.e. sensor node can aggregate data belonging to the same multicast group. Constraint (3) requires that if one path is selected for the group $g$ destined to the destination $d$, the path must also be on the tree adopted by the multicast group $g$.

Constraints (4) and (5) require that number of links on the multicast tree adopted by multicast group $g$ be at least the maximum of $h_g$ and $\left|D_g\right|$. Note that both $h_g$ and $\left|D_g\right|$ are legitimate lower bounds on the number of links on the multicast tree adopted by the multicast group $g$ [10]. From the computational experiments, introducing Constraint (5) will significantly improve the solution quality. Note that $\left|D_g\right|$ and $h_g$ could be calculated in advance, as shown in the **Calculate_**$h_g$ algorithm proposed in [10].

The left hand side term of Constraint (6) calculates the number of paths destined for data source nodes pass through link $l$ for a multicast group. The right hand side term of Constraint (6) is at most $\left|D_g\right|$. When the union of the paths destined for the data source nodes does exist a cycle, and this cycle contains link $l$, then Constraint (6) would not be satisfied since there would be many paths passing through this link. In other words, Constraint (6) is to restrict the union of the paths destined for data source nodes contains a cycle. Constraints (7) and (8) require that any multicast group g selects exactly one path destined for its destination $d$. By enforcing Constraints (6), (7) and (8), the union of the paths shall be a tree.

## 3. Lagrangean Relaxation to DCR Model

The algorithm development is based upon Lagrangean relaxation. In (IP), by introducing Lagrangean multiplier vector $u^1, u^2, u^3$, we dualize Constraints (1), (3) and (6) to obtain the following Lagrangean relaxation problem (LR).

$$Z_D(u^1, u^2, u^3) = \min \sum_{l\in L} a_l C_l + \sum_{l\in L} u_l^1 \left( \sum_{g\in G} y_{gl} - C_l \right) + \sum_{g\in G} \sum_{d\in D_g} \sum_{l\in L} u_{gdl}^2 \left( \sum_{p\in P_{gd}} x_{gpd}\delta_{pl} - y_{gl} \right) + \sum_{g\in G} \sum_{l\in L} u_{gl}^3 \left( \sum_{d\in D_g} \sum_{p\in P_{gd}} x_{gpd}\delta_{pl} - |D_g| y_{gl} \right) \quad \text{(LR)}$$

subject to:

$$C_l \in \{0, 1, 2, 3, ...., |G|\} \qquad \forall l \in L \qquad (9)$$
$$y_{gl} = 0 \text{ or } 1 \qquad \forall g \in G, l \in L \qquad (10)$$
$$\sum_{l\in L} y_{gl} \geq \max\left\{h_g, \left|D_g\right|\right\} \qquad \forall g \in G \qquad (11)$$
$$\sum_{p\in P_{gd}} x_{gpd} = 1 \qquad \forall g \in G, d \in D_g \qquad (12)$$
$$x_{gpd} = 0 \text{ or } 1 \qquad \forall g \in G, d \in D_g, p \in P_{gd}. \qquad (13)$$

We can decompose (LR) into three independent subproblems.

Subproblem 1: for $C_l$

$$\min \sum_{l\in L} (a_l - u_l^1) C_l \qquad \text{(SUB1)}$$

subject to (9).

Subproblem 2: for $y_{gl}$

$$\min \sum_{g\in G} \sum_{l\in L} (u_l^1 - u_{gl}^3 |D_g|) y_{gl} - \sum_{g\in G} \sum_{l\in L} \sum_{d\in D_g} u_{gdl}^2 y_{gl} \qquad \text{(SUB2)}$$

subject to (10) and (11).

Subproblem 3: for $x_{gpd}$

$$\min \sum_{g\in G} \sum_{d\in D_g} \sum_{l\in L} \sum_{p\in P_{gd}} (u_{gdl}^2 + u_{gl}^3) x_{gpd} \qquad \text{(SUB3)}$$

subject to (12) and (13).

(SUB1) can be further decomposed into $|L|$ independent subproblems. For each link $l$,

$$\min(a_l - u_l^1) C_l \qquad \text{(SUB1,1)}$$

subject to:

$$C_l \in \{0, 1, 2, 3, ...., |G|\} \qquad (14)$$

If coefficient of link $l$, $(a_l - u_l^l)$, is negative then set $C_l$ to be $|G|$ otherwise 0. The computational complexity of (SUB1) is $O(1)$ for each link $l$.

(SUB2) can be further decomposed into $|G|$ independent subproblems. For each multicast group $g$,

$$\min \sum_{l\in L} \left( u_l^1 - u_{gl}^3 |D_g| - \sum_{d\in D_g} u_{gdl}^2 \right) y_{gl} \qquad \text{(SUB2.1)}$$

subject to:

$$y_{gl} = 0 \text{ or } 1 \qquad \forall l \in L \qquad (15)$$
$$\sum_{l\in L} y_{gl} \geq \max\left\{h_g, \left|D_g\right|\right\} \qquad (16)$$

By assigning the arc weight of each link $l$ to be $u_l^1 - u_{gl}^3 |D_g| - \sum_{d\in D_g} u_{gdl}^2$, the algorithm proposed in [10] could optimally solve (SUB2.1). The computational complexity of the algorithm is $O\left(|L|(|D_g| + \log |L|)\right)$ for each multicast group $g$.

(SUB3) can be further decomposed into $\sum_{g\in G} \left|D_g\right|$ independent shortest path problems with nonnegative arc weight. For each shortest path problem it can be effectively solved by Dijkstra's algorithm. The computational complexity of Dijkstra's algorithm is $O(|N|^2)$ for each destination of the multicast group.

According to the algorithms proposed above, we could effectively solve the Lagrangean relaxation problem optimally. Based on the weak Lagrangean duality theorem, $Z_D(u^1, u^2, u^3)$ is a lower bound on $Z_{IP}$. We could calculate the tightest lower bound by using the subgradient method [1].

## 4. Getting Primal Feasible Solutions to DCR Model

To obtain the primal feasible solutions to the basic model of data-centric wireless sensor routing problem, solutions to the Lagrangean relaxation (LR) is considered. We propose the following two heuristics to get primal feasible solutions.

The first heuristic is to construct shortest path tree based on the solutions in (SUB3). However, in (SUB3), the union of the shortest path for each data source node may not be a tree since the multiplier $u_{gdl}^2$ is associated with each data source node $d$. In other words, each data source node may have different arc weight on link $l$, this results in the possibility of having cycle for the union of the shortest paths. Therefore, we set the arc weight of link $l$ to be

$(\sum_{d \in D_g} u^2_{dgl})/|D_g| + u^3_{gl} + a_l$ , so that the arc weight for link $l$ is the same for all data source nodes of multicast group g. This ensures that the union of the shortest paths destined to every data source in a multicast group shall be a tree. In order to take account the transmission cost, we also incurs $a_l$ on the arc weight. The computational complexity for first heuristic is $O(|G||N|^2)$.

The basic idea of the second heuristic is GIT. According to [6], GIT is a better heuristic than shortest path tree heuristics. By leveraging on the solutions to the dual problem (LR), we set the arc weight for link $l$ as $a_l + u^3_{gl}$. And then we implement the GIT heuristics to construct the tree. The first term $a_l$ is used to reflect the transmission cost. The second term $u^3_{gl}$ reflects the penalty cost for link $l$ to be a link in a cycle. By incorporating the $a_l + u^3_{gl}$ as the arc weight, we try to achieve minimum transmission cost and the gain from data-centric routing (tree) at the same time. The computational complexity of second heuristic is $O(|N|^2 \times \sum_{g \in G} |D_g|)$.

In the following, we show that complete algorithm (denoted as LGR) to solve (IP).

**Algorithm LGR**
**begin**
  Initialize the Lagrangean multiplier vector $(u^1, u^2, u^3)$ to be all zero vectors;
  **run Calculate**_$h_g$;
  $UB$ := very large number; $LB$ := 0;
  *improve_counter* := 0; *step_size_coefficient* := 2;
  **for** *iteration* := 1 **to** *Max_Iteration_Number* **do**
  **begin**
    **run** subproblem (SUB1);
    **run** subproblem (SUB2);
    **run** subproblem (SUB3);
    calculate $Z_D$;
    **if** $Z_D > LB$ **then** $LB := Z_D$ and *improve_counter* := 0;
    **else** *improve_counter* := *improve_counter* + 1;
    **if** *improve_counter* = *Improve_Threshold* **then**
      *improve_counter* := 0; $\delta := \delta / 2$;
    **run** Primal_Heuristic_Algorithm;
    **if** $ub < UB$ **then** $UB := ub$;
    /* *ub* is the newly computed upper bound. */
    **run** update-step-size;
    **run** update-Lagrangean-multiplier;
  **end**;
**end**;

The computational complexity for algorithm LGR is $O(|N|^2 \times \sum_{g \in G} |D_g| + |L||G| \log |L|)$ for each iteration.

## 5. Computational Experiments to DCR Model

The proposed algorithms to basic model for the data-centric routing problem developed in Sects. 3 and 4 are coded in C and run on a PC with $INTEL^{TM}$ PIII-1.3G. *Max_Iteration_Number* and *Improve_Threshold* are set to 2000 and 50 respectively. The step size coefficient $\delta$ is initialized to be 2 and will be halved when the objective function value of the dual problem is not improved for iterations up to *Improve_Threshold*.

Two source placement models, namely, *event-driven* and *random-source* model are tested. In *random-source* model, non-sink nodes are randomly selected to be data source nodes. Unlink in *event-driven* model, the source nodes are not necessarily clustered. Query-based applications and periodic applications could be classified as the *random-source* model. We construct the network topology for $|N|$=300 nodes which are randomly placed in a $1 \times 1$ square unit area. The power aware transmission cost ($a_l$) is defined as $100 \times$ Euclidean distance if link length does not exceed the transmission radius. In Fig. 3 and Fig. 5, maximum communication radius is configured as 0.125. That is to say $a_l = 100 \times$ Euclidean distance if length of link $l \leq 0.125$, otherwise $a_l = \infty$. In Figs. 3–6, SPT, CNS and GIT are the solution approaches proposed in [6]. Heuristic 1 and heuristic 2 are the solution approaches proposed in Sect. 4. Each plotted point in Figs. 3–6 is a mean value over 5 simulation results.

Figure 3 shows the transmission cost of different number of source nodes in random-source model. We could see that the second heuristics proposed in Sect. 4 outperforms than the other four solution approaches under all different number of source nodes. As the number of data source nodes grows, the improvement ratio is more significant. Figure 5 shows the similar computational results for event-driven model. Figure 4 shows the transmission cost for different communication radii for fixed 10 source nodes in random-source model. Heuristic 2 still outperforms than other approaches. Note that as decreasing the communication radius, the improvement ratio of second heuristic is larger. This occurs because when transmission radius is small, only links with shorter distance could exist. The routing path must have more hops in order to reach destination. Therefore, the advantage resulting from data aggregation will be more significant.

Similar computational results could also be observed

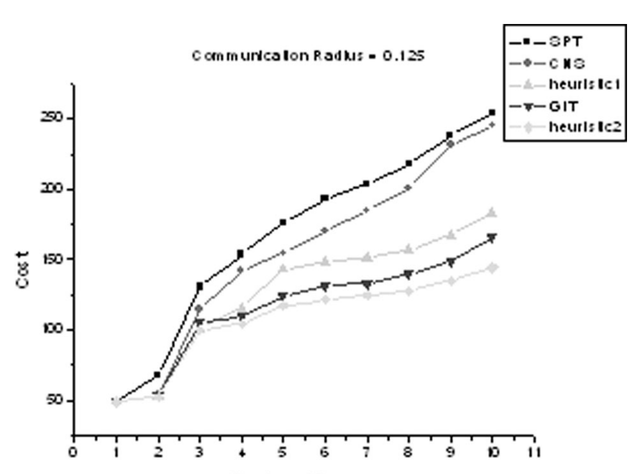



**Fig. 3** Transmission cost vs. number of sources in random-source model.

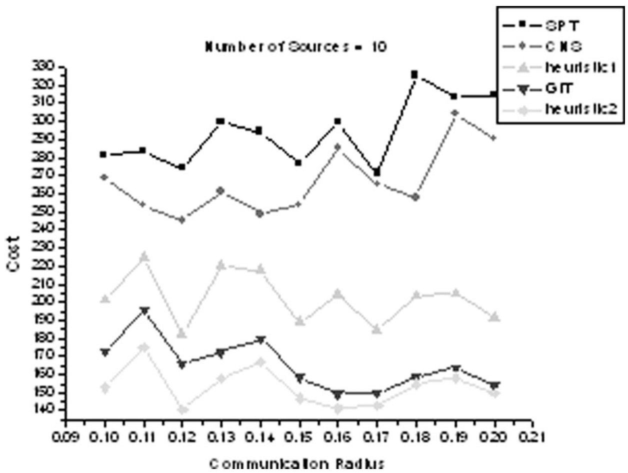



**Fig. 4** Transmission cost vs. communication radius in random-source model.

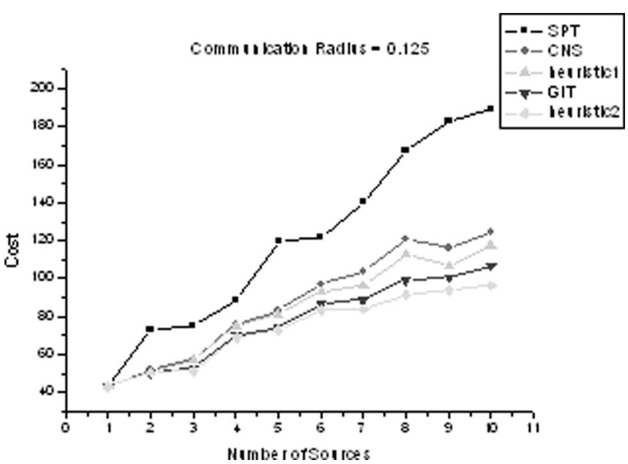



**Fig. 5** Transmission cost vs. number of sources in event-driven model.

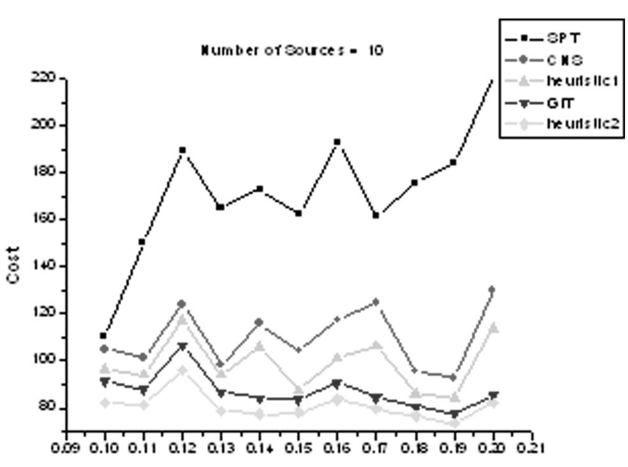



**Fig. 6** Transmission cost vs. communication radius in event-driven model.

**Table 1** Improvement ratio of heuristic 2.

| Improvement Ratio | Fig. 3 | Fig. 4 | Fig. 5 | Fig. 6 |
|---|---|---|---|---|
| SPT | 75 | 110 | 97 | 169 |
| CNS | 71 | 94 | 33 | 58 |
| GIT | 15 | 18 | 11 | 12 |

in Fig. 6 for event-driven model. It is interesting to observe that the improvement ratio in random-source model is often larger than in event-driven model. That is because sources are randomly selected not clustered in random-source model, and the advantages of tree will be more significant. In order to measure how good our heuristic 2 algorithm than the other approaches, we define an improvement ratio which is defined (other approach – heuristic 2)/(heuristic 2) × 100. From Table 1, the improvement ratio of heuristic 2 over SPT, CNS and GIT is up to 169%, 94% and 18% respectively.

## 6. Problem Formulation to EDCR Model

We first show the notations of the EDCR model.

| | |
|---|---|
| $N$ | The set of all sensor nodes |
| $P_{sq}$ | The set of all candidate paths that the data source node $s$ connects to the sink node $q$ |
| $S$ | The set of all data source nodes |
| $h$ | Longest distance of shortest path to reach farthest data source node |
| $\delta_{p(n,k)}$ | The indicator function which is 1 if the link $(n, k)$ is on the path $p$ and otherwise 0 |
| $d_{nk}$ | Euclidean distance between node $n$ and node $k$ |
| $q$ | The data sink node |
| $R_n$ | The set of all possible transmission radii that the node $n$ can adopt |
| $e_n(r_n)$ | Energy consumption function of node $n$, which is a function of nodes transmission radius |

In this EDCR model, we introduce the outgoing link constraint (show in Constraint (20)) to further enforce the tree constraint. In order to model the outgoing link constraint, we model the link $l$ as the node pair $(n, k)$. $n$ is the origin node of the link $l$ and $k$ is the termination node of link $l$. The decision variables are denoted as follows.

$$Z_{IP1} = \min \sum_{n \in N} e_n(r_n) \qquad \text{(IP1)}$$

subject to:

$$\sum_{p \in P_{sq}} x_{sp} \delta_{p(n,k)} \le y_{(n,k)} \qquad \forall n, k \in N, s \in S \qquad (17)$$

$$\sum_{n \in N} \sum_{k \in N} y_{(n,k)} \ge \max\{h, |S|\} \qquad (18)$$

$$\sum_{s \in S} \sum_{p \in P_{sq}} x_{sp} \delta_{p(n,k)} \leq |S| \cdot y_{(n,k)} \quad \forall n, k \in N \quad (19)$$

$$\sum_{k \in N} y_{(n,k)} \leq 1 \quad \forall n \in N \quad (20)$$

$$y_{(n,k)} d_{nk} \leq r_n \quad \forall n, k \in N \quad (21)$$

$$r_n \in R_n \quad \forall n \in N \quad (22)$$

$$y_{(n,k)} = 0 \text{ or } 1 \quad \forall n, k \in N \quad (23)$$

$$\sum_{p \in P_{sq}} x_{sp} = 1 \quad \forall s \in S \quad (24)$$

$$x_{sp} = 0 \text{ or } 1 \quad \forall s \in S, p \in P_{sq}. \quad (25)$$

| | |
|---|---|
| $r_n$ | Transmission radius of the node $n$ |
| $y_{(n,k)}$ | 1 if the link $(n,k)$ is on the tree |
| $x_{sp}$ | 1 if the data source node $s$ uses the path $p$ to reach the sink node $q$ |

The objective function of (IP1) is to minimize total power consumption of the data aggregation tree for transmitting data to the sink node. Constraint (17) requires that if one path $p$ is selected for source node s to reach sink node $q$, the path must also be on the tree. This constraint also enforces that if link $(n,k)$ is on the path $p$ adopted by source node s to reach sink node, then $y_{(n,k)}$ must be 1.

Constraint (18) and (23) require that total number of links on the aggregation tree be at least the maximum of $h$ and the cardinality of $S$. Just as (IP) in Sect. 2, introducing constraint (18) is to improve the solution quality.

Constraint (19) is to restrict the union of the paths destined for data source nodes contains a cycle just as Constraint (6) in (IP). Constraint (20) is an outgoing link constraint. All intermediate nodes on the aggregation tree should have only one outgoing link (e.g. node 4 has two incoming link and only one outgoing link in Fig. 2(a)). Constraint (24) and (25) require that any data source adopts only one path destined for sink node in aggregation tree. By enforcing Constraints (19), (20), (24), and (25) the union of the paths shall be a reverse multicast tree.

Constraint (21) is a transmission radius coverage constraint. This constraint enforces that if link $(n,k)$ is used by aggregation tree, the transmission radius of node $n$ should be large enough in order to cover node $k$. Constraint (22) indicate the set of possible transmission radii for sensor node, which is a discrete and finite set. By enforcing Constraints (21) and (22), we ensure that every link on the aggregation tree is covered within the transmission radius of the origin node of the link.

After presenting the mathematical formulation of the EDCR model, we could summarize the major difference between the DCR and EDCR model. In the DCR model, after the maximum transmission radius is given, the topology of whole network can be constructed. Hence, the data centric aggregation algorithm developed for DCR model could also be applied to wired sensor network when $a_l$ represents the link cost of physical link. On the other hand, in the EDCR model, the transmission radius of sensor node is also a decision variable. In other words, network topology needs to be determined by the transmission radius assignment of the sensor node. Such transmission radius assignment makes EDCR model more general than the DCR model.

## 7. Lagrangean Relaxation to EDCR Model

The algorithm development is based upon Lagrangean relaxation. In (IP1), by introducing Lagrangean multiplier vector $v^1, v^2, v^3$, we dualize Constraints (17), (19) and (21) to obtain the following Lagrangean relaxation problem (LR1).

$$\begin{aligned} Z_{D1}(v^1, v^2, v^3) &= \min \sum_{n \in N} e_n(r_n) \\ &+ \sum_{n \in N} \sum_{k \in N} \sum_{s \in S} v^1_{(n,k)s} \left( \sum_{p \in P_{sq}} x_{sp} \delta_{p(n,k)} - y_{(n,k)} \right) \\ &+ \sum_{n \in N} \sum_{k \in N} v^2_{(n,k)} \left( \sum_{s \in S} \sum_{p \in P_{sq}} x_{sp} \delta_{p(n,k)} - |S| \cdot y_{(n,k)} \right) \\ &+ \sum_{n \in N} \sum_{k \in N} v^3_{(n,k)} (y_{(n,k)} d_{nk} - r_n) \end{aligned} \quad \text{(LR1)}$$

subject to:

$$\sum_{n \in N} \sum_{k \in N} y_{(n,k)} \geq \max\{h, |S|\} \quad (26)$$

$$\sum_{k \in N} y_{(n,k)} \leq 1 \quad \forall n \in N \quad (27)$$

$$r_n \in R_n \quad \forall n \in N \quad (28)$$

$$y_{(n,k)} = 0 \text{ or } 1 \quad \forall n, k \in N \quad (29)$$

$$\sum_{p \in P_{sq}} x_{sp} = 1 \quad \forall s \in S \quad (30)$$

$$x_{sp} = 0 \text{ or } 1 \quad \forall s \in S, p \in P_{sq}. \quad (31)$$

We can decompose (LR1) into three independent subproblems.

Subproblem 4: for $r_n$

$$\min \sum_{n \in N} \left( e_n(r_n) - r_n \sum_{k \in N} v^3_{(n,k)} \right) \quad \text{(SUB4)}$$

subject to (28).

Subproblem 5: for $y_{(n,k)}$

$$\min \sum_{n \in N} \sum_{k \in N} \left( v^3_{(n,k)} d_{nk} - v^2_{(n,k)} |S| - \sum_{s \in S} v^1_{(n,k)s} \right) y_{(n,k)} \quad \text{(SUB5)}$$

subject to (26), (27), and (29).

Subproblem 6: for $x_{sp}$

$$\min \sum_{n\in N}\sum_{k\in N}\sum_{s\in S}\sum_{p\in P_{sq}} (v^1_{(n,k)s} + v^2_{(n,k)})x_{sp}\delta_{p(n,k)} \qquad \text{(SUB6)}$$

subject to (30) and (31).

(SUB4) can be further decomposed into $|N|$ independent subproblems. For each node $n$,

$$\min e_n(r_n) - r_n \sum_{k\in N} v^3_{(n,k)} \qquad \text{(SUB4.1)}$$

subject to:

$$r_n \in R_n \qquad (32)$$

Since $R_n$ is a finite and discrete set. We could examine all possible transmission radii of node $n$ to identify the smallest value of the objective function. The computational complexity of (SUB4) is $O(|R_n|)$ for each node $n$.

The proposed algorithm for solving (SUB5) is described as follows:

**Step 1:** For each link $(n, k)$ compute the coefficient $v^3_{(n,k)}d_{nk} - v^2_{(n,k)}|S| - \sum_{s\in S} v^1_{(n,k)s}$ for each $y_{(n,k)}$.

**Step 2:** For all outgoing links of node $n$, find the smallest coefficient. If the smallest coefficient is negative then set the corresponding $y_{(n,k)}$ to be 1 and the other outgoing links $y_{(n,k)}$ to be 0, otherwise set all outgoing link $y_{(n,k)}$ to be 0. Repeat step 2 for all nodes.

**Step 3:** If the total number of $y_{(n,k)}$ whose value is 1 (denote as $T$) are smaller than $\max\{h, |S|\}$, then identify the nodes that have all its outgoing links $y_{(n,k)} = 0$. From these identified nodes, selected $(\max\{h, |S|\} - T)$ number of these identified nodes whose corresponding smallest coefficients are the smallest. Then, assign the outgoing link $y_{(n,k)} = 1$ with the smallest coefficient for each of these selected nodes.

The computational complexity of above algorithm is $O(|N|^2)$.

(SUB6) can be further decomposed into $|S|$ independent shortest path problems with nonnegative arc weight whose value is $v^1_{(n,k)s} + v^2_{(n,k)}$. For each shortest path problem it can be effectively solved by Dijkstra's algorithm. The computational complexity of Dijkstra's algorithm is $O(|N|^2)$ for each source node.

According to the algorithms proposed above, we could effectively solve the Lagrangean relaxation problem (LR1) optimally. Based on the weak Lagrangean duality theorem, $Z_{D1}(v^1, v^2, v^3)$ is a lower bound on $Z_{IP1}$. We could calculate the tightest lower bound by using the subgradient method [1].

## 8. Getting Primal Feasible Solutions to EDCR Model

To obtain the primal feasible solutions to the extension of data-centric wireless sensor routing problem, solutions to the Lagrangean relaxation (LR1) is considered. We propose the following two heuristics to get primal feasible solutions.

The first heuristic is to construct shortest path tree based on the solutions in (SUB6). However, in (SUB6), the union of the shortest path for each data source node may not be a tree since the multiplier $v^1_{(n,k)s}$ is associated with each data source node s. In other words, each data source node may have different arc weight on link $(n, k)$, this results in the possibility of having cycle for the union of the shortest paths. Therefore, we set the arc weight of link $(n, k)$ to be $\sum_{s\in S} v^1_{(n,k)s}/|S| + v^2_{(n,k)} + (d_{nk}/4)^2$, so that the arc weight for link $(n, k)$ is the same for all data source nodes. This ensures that the union of the shortest paths destined to every data source shall be a tree. In order to take account the transmission power consumption, we also incurs transmission distance $d_{nk}$ between node $n$ and $k$ on the arc weight. After the aggregation tree is determined, the minimum power to cover each link on the tree could be determined. The computational complexity for first heuristic is $O(|N|^2)$.

The principle idea of the second heuristic is also leveraging on GIT. We set the arc weight for link $(n, k)$ as $v^2_{n,k} + (d_{nk}/4)^2$ and then running the GIT algorithm. The idea of divide $d_{nk}$ by 4 is for normalization purpose such that arc weight will not be dominated by $d_{nk}$. The first term $v^2_{(n,k)}$ reflects the penalty cost for link (n,k) to be a link in a cycle. The second term $(d_{nk}/4)^2$ is used to reflect the transmission power consumption. By incorporating the $v^2_{n,k} + (d_{nk}/4)^2$ as the arc weight, we try to achieve minimum transmission cost and the gain from data-centric routing (tree) at the same time. After the aggregation tree is determined, the minimum power to cover each link on the tree could be determined. The computational complexity of second heuristic is $O(|N|^2 \times |S|)$.

In the following, we show that complete algorithm (denoted as LGR1) to solve (IP1).

**Algorithm LGR1**
**begin**
  Initialize the Lagrangean multiplier vector $(v^1, v^2, v^3)$ to be all zero vectors;
  **run Calculate_$h_g$** to determine $h$;
  $UB$ := very large number; $LB$ := 0;
  *improve_counter* := 0; *step_size_coefficient* := 2;
  **for** *iteration* := 1 **to** *Max_Iteration_Number* **do**
  **begin**
    **run** subproblem (SUB4);
    **run** subproblem (SUB5);
    **run** subproblem (SUB6);
    calculate $Z_{D1}$;
    **if** $Z_{D1} > LB$ **then** $LB := Z_{D1}$ and *improve_counter* := 0;
    **else** *improve_counter* := *improve_counter* + 1;
    **if** *improve_counter* = *Improve_Threshold* **then**
      *improve_counter* := 0; $\delta := \delta / 2$;
    **run** Primal_Heuristic_Algorithm;
    if $ub < UB$ then $UB := ub$;
    /* $ub$ is the newly computed upper bound. */
    **run** update-step-size;
    **run** update-Lagrangean-multiplier;

**end**;
**end**;

And the computational complexity for LGR1 is $O(|N|^2 \times |S| + \sum_{n \in N} |R_n|)$ for each iteration.

## 9. Computational Experiments to EDCR Model

The proposed algorithms to EDCR model developed in Sects. 7 and 8 are coded in C and run on a PC with $INTEL^{TM}$ PIII-1.3G. *Max_Iteration_Number* and *Improve_Threshold* are set to 1000 and 25 respectively. The step size coefficient $\delta$ is initialized to be 2 and will be halved when the objective function value of the dual problem is not improved for iterations up to *Improve_Threshold*. The computational time is all within five minutes for all tested cases.

*Random-source* model is tested. In *random-source* model, non-sink nodes are randomly selected to be data source nodes. We construct the network topology for $|N|$=150 nodes which are randomly placed in a $1 \times 1$ square unit area. The transmission power falls as $1/d^n$ [7], $n \geq 2$, where $d$ represents the Euclidean distance. Thus, the cost of power aware function $e_n(r_n)$ is defined as square of $100 \times$ Euclidean distance if link length does not exceed the maximum transmission radius. The set of all possible transmission radii of sensor node $n$ ($R_n$) is configured to begin from 0 to maximum transmission radius. Elements in the radius set are increased by 0.01 successively. In Fig. 7, maximum transmission radius is set to be 0.15. Each plotted point in Figs. 7 and 8 is a mean value over 5 experimental results. In all of experiments we assume that there is only one group in sensor networks. In order to show the solution quality of our proposed algorithm, we implement three algorithms developed in [6] to determine the aggregation tree and then to determine the minimum transmission cost for comparison.

Figure 7 shows the transmission cost of different number of source nodes in random-source model. We could see that the second heuristics proposed in Sect. 8 is superior to the other four solution approaches under all different number of source nodes. In addition, as the number of data source nodes grows the improvement ratio is more significant, which is similar to the DCR model. Figure 8 shows the transmission cost for different maximum transmission radii fixed 8 source nodes in random-source model. The maximum transmission radius is the maximum allowable transmission range that sensor nodes can chose. Heuristic 2 still outperforms than other approaches. Note that as decrease of the transmission radius, the improvement ratio of second heuristic is larger. This result shows the similar results in the DCR model. Another interesting point is that the cost decreases with increasing maximum transmission radius because of the expanded feasible region. However, we can observe that when the maximum transmission radius is increased to a certain point (e.g. 0.17 in Fig. 8), the cost can not be reduced any more. This is because the power consumption cost is defined as the square of transmission radius, and the cost will be increased rapidly when the large transmission radius is chosen. Therefore, even though the maximum allowable transmission radius is increased, we will not be willing to turn on the larger transmission radius.

From Table 2, the improvement ratio of heuristic 2 in EDCR over SPT, CNS and GIT is up to 59%, 49% and 10% respectively. The convergence behavior of our proposed algorithm is shown in Fig. 9. In most of the cases we observe that the upper bound is good enough (within 1% gap in comparison with the final upper bound) at 100th iteration. Thus, when in real-time environment, we could stop the solution procedure at 100th iteration to reduce the computational time by 10 times (from 5 minutes to 30 seconds) with only at most one percent solution quality penalty. Figure 10

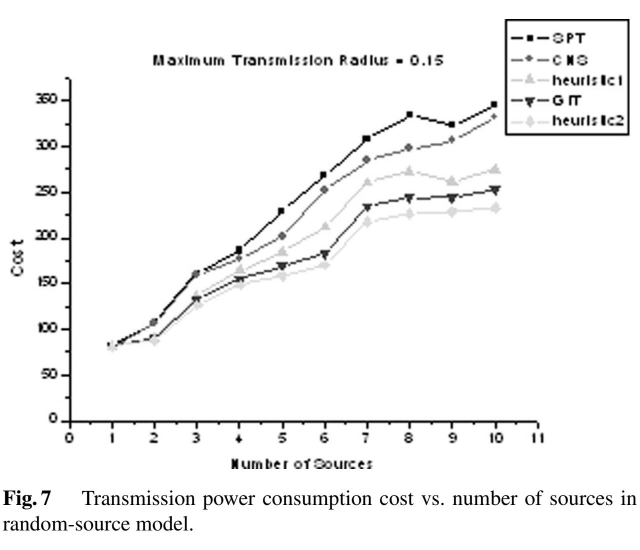



**Fig. 7** Transmission power consumption cost vs. number of sources in random-source model.

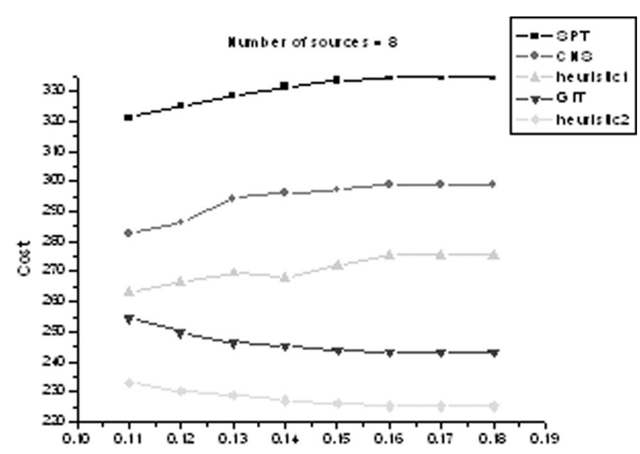



**Fig. 8** Transmission power consumption cost vs. maximum transmission radius in random-source model.

**Table 2** Improvement ratio of heuristic 2.

| Improvement Ratio | Fig. 7 | Fig. 8 |
|---|---|---|
| SPT | 59 | 49 |
| CNS | 49 | 33 |
| GIT | 10 | 10 |

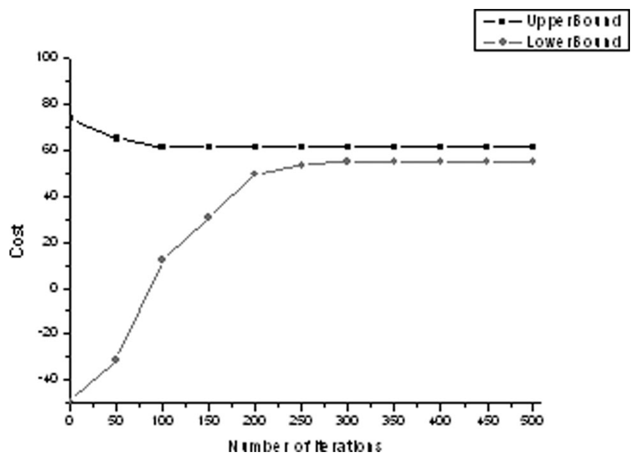



**Fig. 9** Behavior of convergence.

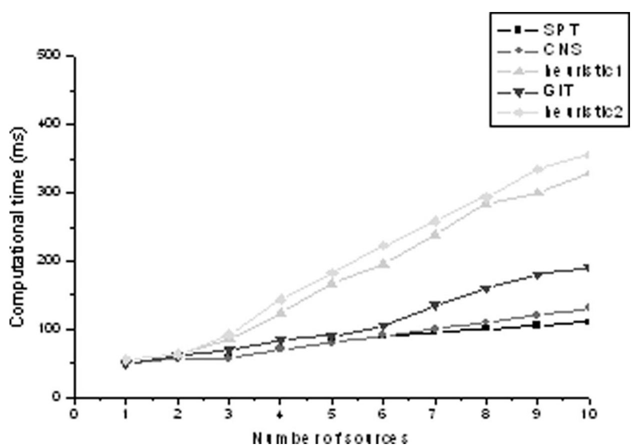



**Fig. 10** Computational time per iteration.

shows the computational time comparison of all algorithms per iteration under different number of sources. Although our proposed algorithms suffer from the slightly longer computational time, we can get better data aggregation tree in terms of transmission cost and energy saving. Furthermore, the improvement ratio of our proposed algorithm will be more significant when the number of source nodes increase. Another interesting point is that when the transmission radius capability of sensor nodes is limited (or equivalently, when sensor nodes are deployed in a large area), the solution quality of our proposed algorithm is even better. To summarize, as compared to existing heuristics, although the disadvantage of our algorithms is the slightly longer computational time, however the advantages of our algorithms are better data-centric aggregation capability and better solution quality particularly in increasing number of source nodes and large sensor node deployment area.

## 10. Conclusion

The data-centric routing could reduce the transmission power for sensor nodes with data aggregation capability in wireless sensor networks. In this paper, we first propose mixed integer and linear mathematical formulation for data centric routing problem. Solution approaches based on Lagrangean relaxation and optimization-based heuristic are proposed to solve this problem. From the computational experiments, the proposed algorithm for DCR problem is superior to the existing approaches (SPT, CNS and GIT [6]) with improvement ratio up to 169%, 94% and 18% respectively.

Besides routing assignment, transmission radius assignment is also considered to address the self-organized property of sensor node. In the second part of this paper, we jointly consider transmission radius assignment and routing assignment in data-centric sensor networks. Lagrangean relaxation techniques in conjunction with optimization-based heuristics are proposed for the EDCR problem. Through experimental results, the proposed algorithm for EDCR model still outperforms all other solution approaches. Besides solution quality, the computational time for the proposed algorithms of DCR problem are all within five minutes in the network topology with randomly generated 300 nodes and 150 nodes. From the solution quality and the computational time, the proposed optimization-based approaches could effectively and efficiently solve the energy-efficient data-centric routing problems in wireless sensor networks.

## Acknowledgement

This work is suported by the National Science Council, Taiwan, under grant number NSC 94-2213-E-128-002.

### References

[1] R.K. Ahuja, T.L. Magnanti, and J.B. Orlin, Networks Flows: Theory, Algorithms, and Applications, Prentice Hall, 1993.
[2] V. Annamalai, S. Gupta, and L. Schwiebert, "On tree-based converge-casting in wireless sensor networks," Proc. IEEE Wireless Communications and Networking Conference, vol.3, pp.1942–1947, 2003.
[3] J. Carle and D. Simplot, "Energy efficient area monitoring by sensor, networks," Computer, vol.37, no.2, pp.40–46, 2004.
[4] M.R. Garey and D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-completeness, Freeman, San Francisco, 1979.
[5] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed fiffusion: A scalable and robust communication paradigm for sensor networks," ACM/IEEE MOBICOM 2000, pp.56–67, Aug. 2000.
[6] B. Krishnamachari, D. Estrin, and S. Wicker, "Modeling datacentric routing in wireless sensor networks," USC Computer Engineering Technical Report CENG 02-14, 2002.
[7] V. Rodoplu and T.H. Meng, "Minimum energy mobile wireless networks," IEEE J. Sel. Areas Commun., vol.17, no.8, pp.1333–1344, Aug. 1999.
[8] S. Singh, M. Woo, and C.S. Raghavendra, "Power-aware routing in mobile ad hoc networks," ACM/IEEE MOBICOM 1998, pp.181–190, 1998.
[9] H.O. Tan and I. Korpeoglu, "Power efficient data gathering and aggregation in wireless sensor networks," ACM SIGMOD Record, vol.32, no.4, pp.66–71, 2003.
[10] H.H. Yen and F.Y.S. Lin, "Near-optimal tree-based access network design," Comput. Commun., vol.28, no.2, pp.236–245, Feb. 2005.

**Hong-Hsu Yen** received his B.S. degree in Industrial Engineering from National Tsing Hua University in 1990, and M.S. degree in Electrical Engineering from National Taiwan University in 1995 and Ph.D. degree in Information Management from National Taiwan University in 2001. He joined the faculty of the Department of Information Management at Shin Hsin University in 2001. His research interests include network planning, network optimization, performance evaluation and QoS routing.

**Frank Yeong-Sung Lin** received his B.S. degree in electrical engineering from the Electrical Engineering Department, National Taiwan University in 1983, and his Ph.D. degree in electrical engineering from the Electrical Engineering Department, University of Southern California in 1991. After graduating from the USC, he joined Telcordia Technologies (formerly Bell Communications Research, abbreviated as Bellcore) in New Jersey, U.S.A., where he was responsible for developing network planning and capacity management algorithms. In 1994, Prof. Lin joined the faculty of the Electronic Engineering Department, National Taiwan University of Science of Technology. Since 1996, he has been with the faculty of the Information Management Department, National Taiwan University. His research interests include network optimization, network planning, performance evaluation, high-speed networks, wireless communications systems and distributed algorithms.

**Shu-Ping Lin** received his B.S. degree in Information Management Department from National Chi Nan University in 2003 and M.S. degree in Information Management from National Taiwan University in 2005. His research interests include network optimization and algorithms.