Efficient Data-Centric Routing in Wireless Sensor Networks

Hong-Hsu Yen
Dept. of Information Management,
Shih Hsin University,
Taipei 116, Taiwan,
hhyen@cc.shu.edu.tw

Frank Yeong-Sung Lin
Dept. of Information Management,
National Taiwan University,
Taipei 106, Taiwan,
yslin@im.ntu.edu.tw

Shu-Ping Lin
Dept. of Information Management,
National Taiwan University,
Taipei 106, Taiwan,
r92725018@ntu.edu.tw

***Abstract*—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 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. The solution approach is based on Lagrangean relaxation in conjunction with the optimization-based heuristics. From the computational experiments, it is shown that the proposed algorithm calculates better solution than other existing heuristics with improvement ratio up to 169% for network with 300 random generated nodes in five minutes of computational time.**

Keywords—Data aggregation, data-centric routing, Lagrangean relaxation, mixed integer linear programming, wireless sensor networks

I. 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 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 from multiple data source nodes to one data sink node. This is a kind of *reverse-multicast* rather than communication between any two pair of nodes in MANETs. Second, since data are collected by multiple sensors there must be some redundancy in the 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 [3, 4].

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 sensor, and the sink node as *data sink*. The application scenario described above is called *event-driven* that 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 then 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. 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 routing in such a way to save energy of sensor. 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. Fig. 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

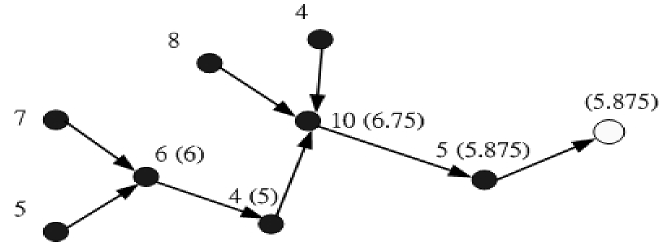

Fig.1. An illustrative example of data aggregation.

0-7803-8938-7/05/$20.00 (C) 2005 IEEE

temperature measurement 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 for Steiner tree problem, which is proven to be the NP-completeness [2]. This calls for the effective and efficient heuristic to solve this problem.

Great deals of existing research have been conducted to address the routing problem in wireless sensor networks. S. Singh [4] 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 Section II), instead of hops which is used in [3], as the link cost.

Krishnamachari [3] devises three interesting suboptimal aggregation heuristics, namely, Shortest Paths Tree (SPT), Center at Nearest Source (CNS), and Greedy Incremental Tree (GIT), respectively. Fig. 2 is a simple illustration of these three heuristics. Note that the transmission cost on each link 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. Fig. 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. Fig. 2(c) shows the final routing assignment by adopting CNS heuristic.

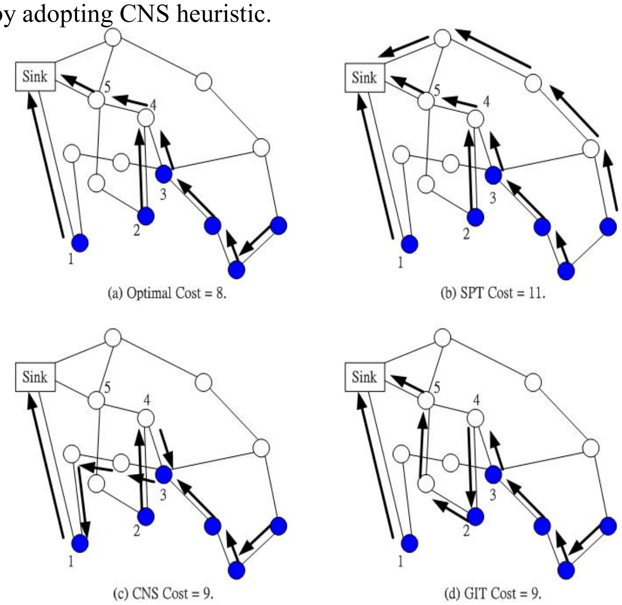

Fig. 2. Simple illustrative example of SPT, CNS and GIT

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 proper 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.

In this paper, we propose an optimization-based heuristics to solve the data-centric routing problems 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. Then Lagrangean relaxation scheme in conjunction with the optimization-based heuristics is proposed to solve this problem. From the computational experiments, the proposed solution approaches outperform the existing heuristics.

The remainder of this paper is organized as follows. In Section II, a MILP formulation of the data-centric wireless sensor networks routing problem is proposed. In Section III, solution approaches based on Lagrangean relaxation are presented. In Section IV heuristics are developed for calculating good primal feasible solution. In Section V, computational results are reported. Finally, Section VI concludes this paper.

II. PROBLEM FORMULATION

A data-centric wireless sensor networks 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 multicast group g
L	The set of all links in the graph
P_{gd}	The set of candidate paths from data source node d to the sink node of multicast group g
h_g	Longest distance to reach farthest data source node for multicast group g
a_l	Unit power aware transmission cost associated with link l
δ_{pl}	The indicator function which is 1 if link l is on 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 multicast group g uses link l and 0 otherwise
x_{gpd}	1 if multicast group g uses path p to reach sink node d and 0 otherwise

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

$$Z_{IP} = \min \sum_{l\in L} a_l C_l \quad \text{(IP)}$$

subject to:

$$\sum_{g\in G} y_{gl} \le C_l \quad \forall l \in L \quad (1)$$

$$C_l \in \{0,1,2,3,....,|G|\} \quad \forall l \in L \quad (2)$$

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

$$y_{gl} = 0 \text{ or } 1 \quad \forall g\in G, l\in L \quad (4)$$

$$\sum_{l\in L} y_{gl} \ge \max\{h_g, |D_g|\} \quad \forall g\in G \quad (5)$$

$$\sum_{d\in D_g}\sum_{p\in P_{gd}} x_{gpd}\delta_{pl} \le |D_g| y_{gl} \quad \forall g\in G, l\in L \quad (6)$$

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

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

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 group g destined to destination d, the path must also be on the tree adopted by 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 the cardinality of D_g. Note that both h_g and D_g are legitimate lower bounds on the number of links on the multicast tree adopted by multicast group g [5]. From the computational experiments, introducing Constraint (5) will significantly improve the solution quality. Note that $|D_g|$ and h_g could be calculated in advance, as shown in the **Calculate_h_g** algorithm proposed in [5].

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 $|D_g|$. 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.

III. LAGRANGEAN RELAXATION

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).

$$Z_{D1}(u^1,u^2,u^3)=\min \sum_{l\in L} a_l C_l + \sum_{l\in L} u_l^1 (\sum_{g\in G} y_{gl} - C_l) + \sum_{g\in G}\sum_{d\in D_g}\sum_{l\in L} u_{gdl}^2 (\sum_{p\in P_{gd}} x_{gpd}\delta_{pl} - y_{gl}) + \sum_{g\in G}\sum_{l\in L} u_{gl}^3 (\sum_{d\in D_g}\sum_{p\in P_{gd}} x_{gpd}\delta_{pl} - |D_g| y_{gl}) \quad \text{(LR)}$$

subject to:

$$C_l \in \{0,1,2,3,....,|G|\} \quad \forall l\in L \quad (9)$$

$$y_{gl} = 0 \text{ or } 1 \quad \forall g\in G, l\in L \quad (10)$$

$$\sum_{l\in L} y_{gl} \ge \max\{h_g, |D_g|\} \quad \forall g\in G \quad (11)$$

$$\sum_{p\in P_{gd}} x_{gpd} = 1 \quad \forall g\in G, d\in D_g \quad (12)$$

$$x_{gpd} = 0 \text{ or } 1 \quad \forall g\in G, d\in D_g, p\in P_{gd} . \quad (13)$$

We can decompose (LR) into three independent subproblems.

Subproblem 1: for C_l

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

subject to (9).

Subproblem 2: for y_{gl}

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

subject to (10) and (11).

Subproblem 3: for x_{gpd}

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

subject to (12) and (13).

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

$$\min (a_l - u_l^1) C_l \quad \text{(SUB1.1)}$$

subject to:

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

If coefficient of link l $(a_l - u_l^1)$ 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,

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

subject to:

$$y_{gl} = 0 \ or \ 1 \qquad \forall l \in L \qquad (15)$$

$$\sum_{l \in L} y_{gl} \geq \max\left\{h_g, \left|D_g\right|\right\}. \qquad (16)$$

By assigning the arc weight of each link l to be $u_l^1 - u_{gl}^3 \mid D_g \mid - \sum_{d \in D_g} u_{gdl}^2$, the algorithm proposed in [5] could optimally solve (SUB2.1). The computational complexity of the algorithm is $O(\mid L \mid (\mid D_g \mid + \log \mid L \mid))$ 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 solve 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, $Z_{D1}(u^1,u^2,u^3)$ is a lower bound on Z_{IP}. We could calculate the tightest lower bound by using the subgradient method [1].

IV. GETTING PRIMAL FEASIBLE SOLUTIONS

To obtain the primal feasible solutions to the data-centric wireless sensor routing problems, 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 $(\sum_{d \in D_g} u_{gdl}^2)/|D_g| + u_{gl}^3 + a_l$, 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 [3], 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 $a_l + u_{gl}^3$. 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_{gl}^3 reflects the penalty cost for link l to be a link in a cycle. By incorporating the $a_l + u_{gl}^3$ 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 $O(|N|^2 \times \sum_{g \in G} \left|D_g\right|)$.

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 $O(\mid N \mid^2 \times \sum_{g \in G} \left|D_g\right| + \mid L \mid\mid G \mid \log \mid L \mid)$ for each iteration.

V. COMPUTATIONAL EXPERIMENTS

The proposed algorithms for the data-centric routing problem developed in Sections III and IV are coded in C and run on a PC with INTEL™ PIII-1.3G. *Max_Iteration_Number* and *Improve_Threshold* are set to 2000 and 50 respectively. The step size coefficient δ 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 1×1 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, 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 Fig. 3-6, SPT, CNS and GIT are the solution approaches proposed in [3]. Heuristic 1 and heuristic 2 are the solution approaches proposed in Section IV. Each plotted point in Fig. 3-6 is a mean value over 5 simulation results.

Fig. 3 shows the transmission cost of different number of source nodes in random-source model. We could see that the second heuristics proposed in Section IV 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. Fig. 5 shows the similar computational results for event-driven model. Fig. 4 shows the transmission cost for different communication radius 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 the communication 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 in Fig. 6 for event-driven model.

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 I, the improvement ratio of heuristic 2 over SPT, CNS and GIT is up to 169%, 94% and 18% respectively.

VI. 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 propose a rigorous mixed integer and linear mathematical formulation for data centric routing problem. Novel solution approaches based on Lagrangean relaxation and optimization-based heuristic are proposed to solve this problem. From the computational experiments, the proposed algorithm is superior to the existing approaches (SPT, CNS and GIT [3]) with improvement ratio up to 169%, 94% and 18% respectively. Besides solution quality, the computational time for the proposed algorithms is all within five minutes in the network topology with randomly generated 300 nodes. From the solution quality and the computational time, the proposed optimization-based approaches effectively and efficiently solve the data-centric routing problems in wireless sensor networks.

REFERENCES

[1] R. K. Ahuja, T. L. Magnanti and J. B. Orlin, “Networks Flows—Theory, Algorithms, and Applications”, *Prentice Hall*, 1993.

[2] M. R. Garey and D. S. Johnson, “Computers and Intractability: A Guide to the Theory of NP-completeness”, *Freeman*, San Francisco, 1979.

[3] B. Krishnamachari, D. Estrin, and S.Wicker, "Modeling Data-Centric Routing in Wireless Sensor Networks." *IEEE INFOCOM* 2002.

[4] S. Singh, M. Woo, and C. S. Raghavendra, “Power-Aware Routing in Mobile Ad Hoc Networks”, *ACM/IEEE MOBICOM* 1998.

[5] H. H. Yen and F. Y. S. Lin, “Near-optimal Tree-based Access Network Design”, *Computer Communications*, Vol. 28/2, pp. 236-245, Feb. 2005.

TABLE I. 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

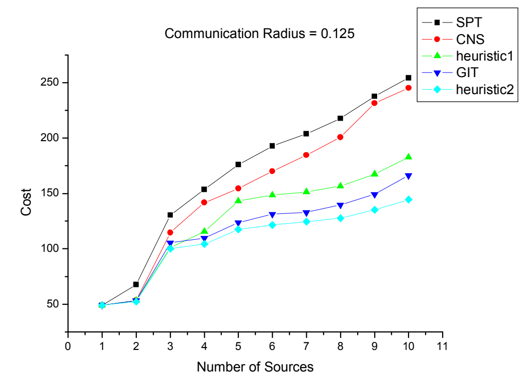

Fig. 3. Transmission cost V.S. number of sources in random-source model

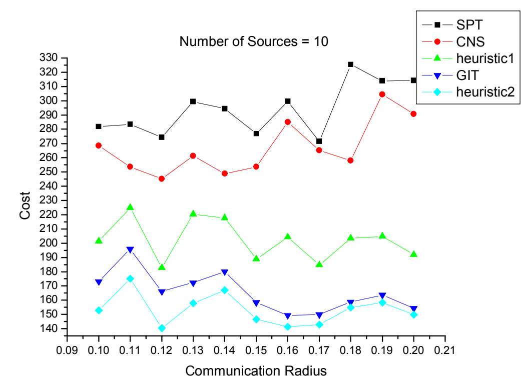

Fig. 4. Transmission cost V.S. communication radius in random-source model

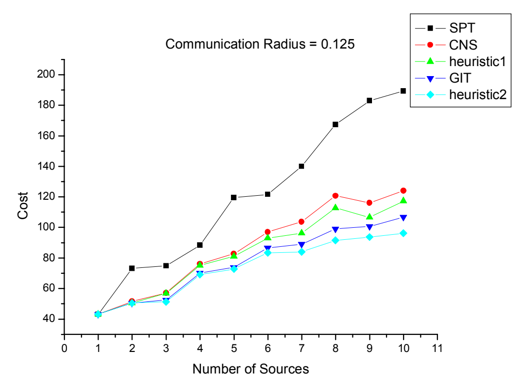

Fig. 5. Transmission cost V.S. number of sources in event-driven model

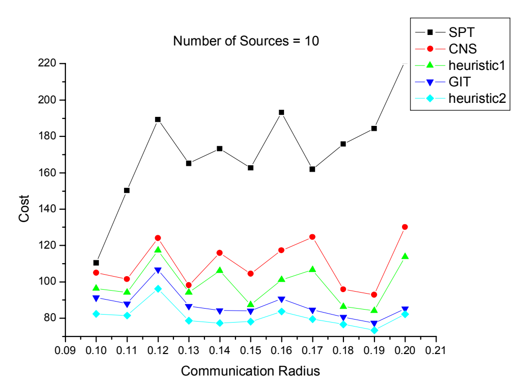

Fig. 6. Transmission cost V.S. communication radius in event-driven model