MAC Aware Energy-Efficient Data-Centric Routing in Wireless Sensor Networks

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

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

Shu-Ping Lin and Yean-Fu Wen
Dept. of Information Management,
National Taiwan University,
Taipei 106, Taiwan,
r92725018@ntu.edu.tw, d89002@im.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. However, the penalty from data retransmissions due to collision could jeopardize the advantages from data aggregation. In this paper, for the first time, we consider the energy consumption tradeoffs between the data aggregation and retransmission in wireless sensor network. By using the CSMA-CA MAC protocol, the retransmission energy consumption function is well formulated. We propose a rigorous non-linear mathematical formulation, where the objective function is to minimize the total energy consumption of data transmission subject to data aggregation tree and data retransmission. 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 algorithms could construct more energy efficient data aggregation tree with MAC layer retransmission mechanism than existing data centric algorithms up to 93%.**

Keywords—Data aggregation, MAC aware energy-efficient data-centric routing, retransmission, Lagrangean relaxation, wireless sensor networks

I. INTRODUCTION

The wireless sensor networks (**WSN**) probe and collect environmental information, such as temperature, atmospheric pressure and irradiation to provide ubiquitous sensing, computing and communication capabilities. WSN has two important and interesting characteristics. First, typical communication mode in WSN 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, if specific event happens, data are collected by multiple sensors and sent back to the sink node. Intermediate node that is on the reverse-multicast path could receive multiple data from the data source nodes. In order not to transmit redundant and useless data back to the data source node, intermediate node along the reverse-multicast path should collect and process data before transmission to save energy. Otherwise, it would results in disconnected network with rapidly energy depletion of sensors. This kind of *Data aggregation* capability has been put forward as a particularly useful function for routing in terms of energy consumption in WSN [2]. In addition to redundancy suppression, other aggregation function could be MAX, MIN, or SUM. For example, in Fig. 1, node n_1, n_2, and n_3 are the data source nodes that probe the temperature (each with 60, 65 and 63) and sent back the MAX temperature back to sink node. If node *S* could aggregate (i.e. MAX = 65) these data before sending back the sink node, the total number of transmission times for node *S* could be reduced from 3 to 1.

From the routing path point of view, interestingly, data are routed along *reverse multicast tree* where multiple data sources transmit information back to the sink node [6]. 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*. In data-centric routing, the key issue is how to construct the reverse multicast tree in such a way to save the total energy consumption. Most of the existing research literatures construct the tree by only taking consideration of data aggregation aspect [2][6]. However, there is one more issue that is important to the construction of data aggregation tree, *MAC layer retransmission* issue.

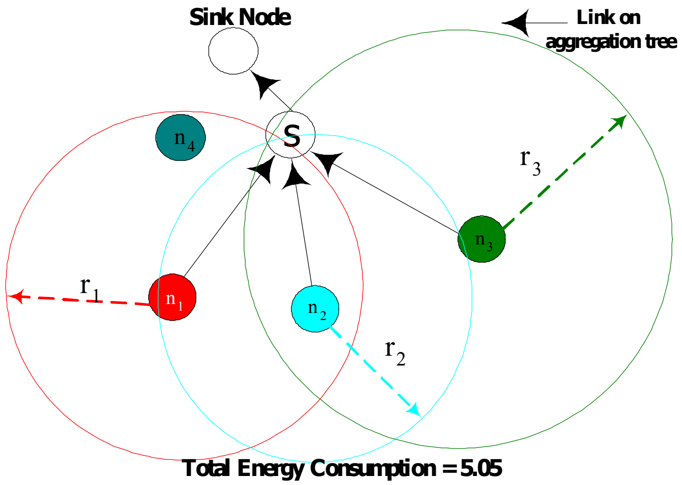

Figure 1. Tradeoff between data aggregation and data retransmissions (without considering data collision due to interference)

In WSN, sensor nodes that are within each other's transmission radius try to transmit simultaneously would result in *collision.* When collision occurs, *retransmission* is required to ensure the data is successfully received. Data retransmission times are affected by the total number of sensor nodes whose transmission radius covers the receiver. In other words, the more flows that the non-leaf node on the aggregation tree are aggregated, the higher probability that the sender will incur data retransmission. Unfortunately, retransmission, which incurs extra energy consumption, will jeopardize the advantage from data aggregation. In other words, constructing good data aggregation tree should consider the tradeoff between data aggregation issue and MAC layer retransmission issue.

1-4244-0355-3/06/$20.00 (c) 2006 IEEE

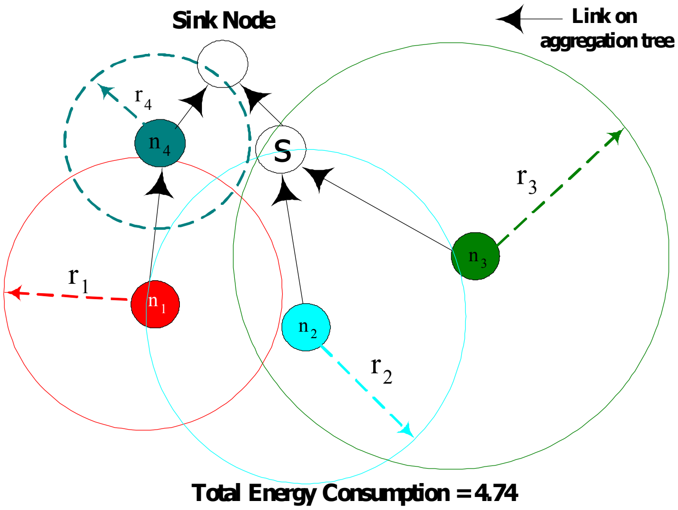

Figure 2. Tradeoff between data aggregation and data retransmissions (with considering collision due to interference)

Fig. 1 and Fig. 2 show the tradeoff between the data aggregation and retransmission. Node n_1, n_2, and n_3 are the data source nodes in Fig. 1 and Fig. 2. Without considering the data collision effect, the optimal aggregation tree is shown in Fig. 1. However, the more data an intermediate node aggregates, the greater the number of collisions that will occur at intermediate node, which results in excess energy consumption. Node *S*, which is the receiver of three children nodes, will suffer severe collisions resulting in more retransmission times. With considering the collision effect, more energy efficient data aggregation tree is shown in Fig. 2. In Fig. 2, by reducing the transmission radius of node n_1 and change its routing assignment to node n_4, the total energy consumption could be reduced. Although we have extra energy consumption at node n_4, but there are only two children nodes at node *S* such that the retransmission times could be significantly reduced. Hence, the energy consumption associated with the MAC aware collision effect should be carefully addressed in wireless sensor networks. Note that the energy consumption function in Fig. 1 and Fig. 2 are calculated by objective function (IP) described in Section II and the function of average retransmission times is given at equation (0) in Section II.

Existing researches have been conducted to address the data centric routing problem in WSN. In [2], they devise three interesting suboptimal aggregation heuristics, Shortest Paths Tree (SPT), Center at Nearest Source (CNS), and Greedy Incremental Tree (GIT) for data centric routing problem. In [6], mathematical formulation for data centric problem in WSN is well formulated and an optimization-based heuristic is then proposed to tackle the problem. In [5], they address latency issue in constructing the minimum energy aggregation tree. They propose the CCA algorithm with the basic idea of balance tree to minimize the energy and latency at the same time.

Several papers have discussed about MAC layer transmission protocol in ad-hoc or sensor networks [7—9]. X. H. Lin [9] enhances the standard IEEE 802.11 MAC protocol by improving the handshaking and power control mechanisms. W. Ye [7][8] reviews several MAC protocol and discusses design trade-offs on energy efficiency and data transmission. W. Ye proposes S-MAC protocol to fit the energy-efficient requirement for sensor networks. It is also a variation of CSMA-like protocol and needs some extra messages for transmitting data. To the best of our knowledge, there is no literature address the MAC aware energy efficient data centric routing algorithm in WSN.

In this paper, we discuss the impacts of retransmission on aggregating data, and propose MAC aware energy efficient data centric algorithm by considering the tradeoff between data aggregating benefits and data retransmission costs in WSN. We propose an optimization-based heuristics to solve the MAC aware energy-efficient data-centric routing problems (**MAC-DCR**) based on CSMA-based protocol in WSN. The problem is first formulated as a nonlinear programming problem where the objective function is to minimize total energy consumption from data transmission and retransmission. 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 conventional non MAC aware data centric heuristics.

The remainder of this paper is organized as follows. In Section II, mathematical formulation of the MAC-DCR in WSN 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

An MAC-DCR in WSN 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.

N	The set of all sensor nodes
P_{sq}	The set of all candidate paths that the data source node *s* connect 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
M	An arbitrary large number
$\delta_{p(n,k)}$	The indicator function which is 1 if the link from node *n* to node *k* is on the path *p* and 0 otherwise
d_{nk}	Euclidean distance between the node *n* and the node *k*
t_{data}	Transmission time for transmitting a data packet
RTS	Transmission time for RTS frame
$SIFS$	Short inter-frame space time
θ	Maximum propagation delay for transmitting data packet
λ	Packet arrival rate
q	The sink node
R_n	The set of all possible transmission radii that the node *n* can adopt, this is a discrete set
$e_n(r_n)$	Energy consumption function of node *n* per unit time, which is a function of sensor's transmission radius
T	The largest number of retransmission times

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.

The decision variables are denoted as follows.

x_{sp}	1 if the data source node s uses the path p to reach the sink node q and 0 otherwise
$y_{(n,k)}$	1 if the link from node n to node k is on the tree and 0 otherwise
r_n	Transmission radius of the node n
z_{nk}	1 if the node k is covered within transmission radius of the node n and 0 otherwise
c_{nk}	Retransmission times of the node n to transmit data to the node k

The analysis of retransmission is conducted as follows. First of all, we assume that each sensor node is equipped with a CSMA-CA compatible transceiver. Based on the analysis in [4], we derive the mean retransmission times of a sender. We assume that each transmission conforms to Geometric distribution and each sensor node generates data packets that follow Poisson distribution with a certain rate, λ. Successful transmission of data from a sender to a receiver is influenced by the number of senders whose transmission radius covers the receiver. By considering the receiver side collisions in terms of the communication radius of sensor nodes, the hidden terminal problem is also implicitly contemplated. In CSMA/CA protocol, when sender want to transmit packet to the receiver, it will first issue the RTS control frame and then waits for the CTS frame from the receiver to make sure the contention is success or not [4]. By considering the turnaround time (which is 2θ) between sender and receiver, the overall contention period is $RTS + SIFS + 2\theta$. Then the average retransmission times from node n to node k is as follows:

$$\textit{Average Retransmission Times}_{(n,k)} = \frac{1}{p_{success(n,k)}} = \frac{1}{e^{-\lambda(RTS+SIFS+2\theta)\sum_{j\in N} z_{jk}}}. \tag{0}$$

The meaning of (0) is the mean value of the Geometric distribution where the successful transmission probability, say $p_{success}$, is that no data transmission is occurring at any node whose transmission radius covers receiver node k within the contention period ($RTS+SIFS+2\theta$). $\sum_{j\in N} z_{jk}$ calculates the total number of senders whose transmission radius covers the node k.

The MAC-DCR in WSN is then formulated as the following combinatorial optimization problem (IP).

$$Z_{IP} = \min \sum_{n\in N} \left(t_{data} + (RTS \sum_{k\in N} c_{nk}) \right) \cdot e_n(r_n) \tag{IP}$$

subject to:

$$\sum_{p\in P_{sq}} x_{sp}\delta_{p(n,k)} \leq y_{(n,k)} \qquad n, k\in N, s\in S \tag{1}$$

$$\sum_{n\in N}\sum_{k\in N} y_{(n,k)} \geq \max\{h, |S|\} \tag{2}$$

$$\sum_{s\in S}\sum_{p\in P_{sq}} x_{sp}\delta_{p(n,k)} \leq |S| \cdot y_{(n,k)} \qquad n,k\in N \tag{3}$$

$$\sum_{p\in P_{sq}} x_{sp} = 1 \qquad s\in S \tag{4}$$

$$\sum_{k\in N} y_{(n,k)} \leq 1 \qquad n\in N \tag{5}$$

$$\frac{r_n - d_{nk}}{M} \leq z_{nk} \qquad n,k\in N \tag{6}$$

$$z_{nk} d_{nk} \leq r_n \qquad n,k\in N \tag{7}$$

$$y_{(n,k)} \leq z_{nk} \qquad n,k\in N \tag{8}$$

$$c_{nk} \geq \frac{e^{-(1-y_{(n,k)})M}}{e^{-\lambda(RTS+SIFS+2\theta)\sum_{j\in N} z_{jk}}} \qquad n,k\in N \tag{9}$$

$$x_{sp} = 0 \text{ or } 1 \qquad \forall s\in S, p\in P_{sq} \tag{10}$$

$$y_{(n,k)} = 0 \text{ or } 1 \qquad n,k\in N \tag{11}$$

$$z_{nk} = 0 \text{ or } 1 \qquad n,k\in N \tag{12}$$

$$r_n \in R_n \qquad n\in N \tag{13}$$

$$r_n \neq 0 \qquad n\in S \tag{14}$$

$$c_{nk} \in \{0, 1, 2, ..., T\} \qquad n,k\in N. \tag{15}$$

The objective function of (IP) is to minimize total energy consumption from data transmission and retransmission. Constraint (1) requires that if the path p is selected for the source node s to reach the sink node q, the path must be on the tree. This constraint also enforces that if the link (n, k) is on the path p adopted by the source node s to reach the sink node, then $y_{(n,k)}$ must be 1. Constraint (2) and (11) require that total number of links on the aggregation tree is at least the maximum of h and the cardinality of S. Note that both h and $|S|$ are legitimate lower bound on the total number of links on an aggregation tree and they could be calculated in advance [3]. Introducing constraint (2) will significantly improve the solution quality. The left-hand term of constraint (3) calculates the number of paths, which are destined for the sink node and passing through the link (n, k) on aggregation tree. The right-hand term of constraint (3) is at most $|S|$. When the union of the paths destined for the sink node does exist a cycle, and this cycle contains link l, then constraint (3) would not be satisfied since there would be many paths pass through this link. In other words, constraint (3) is to enforce that the union of the paths does not contain a cycle [6]. Constraints (4) and (10) require that any data source adopts only one routing path destined for sink node. Constraint (5) is the outgoing link constraint. All intermediate nodes on the aggregation tree should have only one outgoing link. For example, in Fig. 1, node S has only one outgoing link to the sink node. Constraints (3), (4), (5), and (10) enforce that the union of all routing paths shall be a tree.

Constraints (6) and (7) specify the transmission radius coverage constraints. If $r_n \geq d_{nk}$, z_{nk} should be equal to 1 and 0 otherwise. Using z_{nk} we can calculate the total number of sensor nodes whose transmission radius covers sensor node k, and the total number of sensor nodes covered by transmission radius of sensor node n. Constraint (8) is a necessary constraint that relates decision variable $y_{(n,k)}$ to z_{nk}. If $y_{(n,k)}$ equals to 1 then z_{nk} also must be 1.

Constraint (9) calculates the retransmission times of the node n to transmit data to the node k. Note that only the sensor nodes on the aggregation tree need to calculate the retransmission times. Therefore, when $y_{(n,k)} = 1$, the right side of constraint (9) is the same as Equation (0) to enforce the retransmission times should be at least the average retransmission times. When $y_{(n,k)} = 0$, the right side of constraint (9) is zero, which implies no retransmission times constraint. Constraint (13) restricts that the set of possible transmission radii that node *n* can adopt is a discrete and finite set. Constraint (14) enforces that each data source node should turn on its transmission radius. The transmission radius of each source node can not be 0. Constraint (15) is an integer constraint of retransmission times.

We take natural logarithm on both sides for constraint (9) for applying the Lagrangean relaxation schemes,

$$\ln(c_{nk}) \ge \ln\left(\frac{e^{-(1-y_{(n,k)})M}}{e^{-\lambda(RTS+SIFS+2\theta)\sum_{j\in N} z_{jk}}}\right)$$

$$\Rightarrow \ \ln(c_{nk}) \ge \lambda(RTS+SIFS+2\theta)\sum_{j\in N} z_{jk} - M + My_{(n,k)}$$

III. LAGRANGEAN RELAXATION

The algorithm development is based upon Lagrangean relaxation. In (IP), by introducing Lagrangean multiplier vector u^1, u^2, u^3, u^4, u^5, and u^6, we dualize Constraints (1), (3), (6), (7), (8), and (9) to obtain the following Lagrangean relaxation problem (LR).

$$Z_D = \min \sum_{n\in N}\left(\mathrm{t}_{\mathrm{data}} + (RTS\cdot \sum_{k\in N} c_{nk})\right)\cdot e_n(r_n) + \sum_{n\in N}\sum_{k\in N}\sum_{s\in S} u^1_{nks}\left(\sum_{p\in P_{sq}} x_{sp}\delta_{p(n,k)} - y_{(n,k)}\right) + \sum_{n\in N}\sum_{k\in N} u^2_{nk}\left(\sum_{s\in S}\sum_{p\in P_{sq}} x_{sp}\delta_{p(n,k)} - |S|\, y_{(n,k)}\right) + \sum_{n\in N}\sum_{k\in N} u^3_{nk}(r_n - d_{nk} - Mz_{nk}) + $$
$$+ \sum_{n\in N}\sum_{k\in N} u^4_{nk}(z_{nk}d_{nk} - r_n) + \sum_{n\in N}\sum_{k\in N} u^5_{nk}(y_{(n,k)} - z_{nk}) + \sum_{n\in N}\sum_{k\in N} u^6_{nk}\Big(\lambda(RTS+SIFS+2\theta)\sum_{j\in N} z_{jk} - M(1 - y_{(n,k)}) - \ln(c_{nk})\Big) \qquad \text{(LR)}$$

subject to:

$$\sum_{n\in N}\sum_{k\in N} y_{(n,k)} \ge \max\{h, |\mathrm{S}|\} \qquad (16)$$

$$\sum_{p\in P_{sq}} x_{sp} = 1 \qquad s\in S \qquad (17)$$

$$\sum_{k\in N} y_{(n,k)} \le 1 \qquad n\in N \qquad (18)$$

$$x_{sp} = 0 \text{ or } 1 \qquad \forall s\in S, p\in P_{sq} \qquad (19)$$

$$y_{(n,k)} = 0 \text{ or } 1 \qquad n,k\in N \qquad (20)$$

$$z_{nk} = 0 \text{ or } 1 \qquad n,k\in N \qquad (21)$$

$$r_n \in R_n \qquad n\in N \qquad (22)$$

$$r_n \ne 0 \qquad n\in S \qquad (23)$$

$$c_{nk} \in \{0, 1, 2, \ldots, T\} \qquad n,k\in N. \qquad (24)$$

We can decompose (LR) into four independent subproblems.

Subproblem 1: for $y_{(n,k)}$

$$\min \sum_{n\in N}\sum_{k\in N}(u^5_{nk} + u^6_{nk}M - u^2_{nk}|S| - \sum_{s\in S} u^1_{nks})y_{(n,k)} \qquad \text{(SUB1)}$$

subject to (16), (18) and (20).

Subproblem 2: for x_{sp}

$$\min \sum_{n\in N}\sum_{k\in N}\sum_{s\in S}\sum_{p\in P_{sq}}(u^1_{nks} + u^2_{nk})x_{sp}\delta_{p(n,k)} \qquad \text{(SUB2)}$$

subject to (17) and (19).

Subproblem 3: for r_n and c_{nk}

$$\min \sum_{n\in N} e_n(r_n)\cdot t_{data} + RTS\sum_{n\in N}\sum_{k\in N} e_n(r_n)\cdot c_{nk} + \sum_{n\in N}\sum_{k\in N}(u^3_{nk} - u^4_{nk})r_n - \sum_{n\in N}\sum_{k\in N} u^6_{nk}\ln(c_{nk}) \qquad \text{(SUB3)}$$

subject to (22), (23) and (24).

Subproblem 4: for z_{nk}

$$\min \sum_{n\in N}\sum_{k\in N}\left(u^4_{nk}d_{nk} - u^3_{nk}M - u^5_{nk} + \lambda(RTS+CTS+2\theta)\sum_{j\in N} u^6_{jk}\right)z_{nk} \qquad \text{(SUB4)}$$

subject to (21).

The proposed algorithm for solving (SUB1) is described as follows.

Step1. For each link (*n*,*k*) compute the coefficient $u^5_{nk} + u^6_{nk}M - u^2_{nk}|S| - \sum_{s\in S} u^1_{nk}$ for each $y_{(n,k)}$.

Step2. 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.

Step3. If the total number of $y_{(n,k)}$ whose value is 1 (denoted as τ) are smaller than $max\{h, |\mathrm{S}|\}$, then first let each $y_{(n,k)}$ whose corresponding coefficient is negative be 1. Second, assign the ($max\{h, |\mathrm{S}|\} - \tau$) number of $y_{(n,k)}$ to be 1 whose corresponding coefficients are the smallest positive values. Third, let the remaining $y_{(n,k)}$ be 0.

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

(SUB2) can be further decomposed into |*S*| independent shortest path problems with nonnegative arc weight whose value is $u^1_{nks} + u^2_{nk}$. 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 data source node.

(SUB3) can be optimally solved by exhaustively searching the combination of radius r_n and c_{nk}. The computational complexity of (SUB3) is $O(|R_n| \times T)$ for each node *n*.

(SUB4) is an easy problem to be solved. If the corresponding coefficient $u^4_{nk}d_{nk} - u^3_{nk}M - u^5_{nk} + \lambda(RTS + CTS + 2\theta)\sum_{j\in N} u^6_{jk}$ of link (*n*, *k*) is negative then set z_{nk} to be 1, otherwise 0. The computational complexity of (SUB4) is $O(1)$ for each link (*n*, *k*).

According to the algorithms proposed above, we could effectively solve the Lagrangean relaxation problem optimally. Based on the weak Lagrangean duality theorem, $Z_D(u^1,u^2,u^3,u^4,u^5,u^6)$ 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 for a data aggregation tree with MAC aware energy efficient data centric routing, we consider solutions to the Lagrangean relaxation (LR) problem. When the routing path, x_{sp}, for each source node, *s*, is determined, other decision variables (r_n and $y_{(n,k)}$) can be calculated and the total energy consumption of the data aggregation tree can be obtained. However, the solution of

(SUB2), it may violate the tree constraint. Thus, we propose a drop heuristic to eliminate those links that form the cycle on the tree.

The complete algorithm (denote as LGR-Primal) for the getting primal feasible solutions is depicted as follows:

1. Based on the solutions of (SUB2) we can get the set of decision variables, x_{sp}, from which we can decide which link, $y_{(n,k)}$, is used on the routing path by source s. After determining $y_{(n,k)}$, if $y_{(n,k)}$ is 1, we set the arc weight on the corresponding link to be $\frac{\sum_{s\in S} u^1_{nks}}{|S|} + u^2_{nk}$; otherwise, we set the arc weight to be infinity. Repeat this step for every data source node.
2. According to the arc weight calculated in Step 1, we sort the links from smallest to largest arc weight.
3. We sequentially examine all links from the link with the largest arc weight to the smallest, but we ignore the links with infinity costs. We remove each link say link (n, k) from the routing path and check whether every source node still has a routing path to the sink node. If any source node is unable to reach the sink node after removing link (n, k), we restore link (n, k) onto the routing path. If every source nodestill has a routing path to reach the sink node, we remove (n, k) and investigate the next link until all the links used by the union of routing path x_{sp} have been examined.

After executing the step 3 of LGR-Primal, we get a data aggregation tree without any cycles. The computational complexity of the LGR-Primal is $O(|S||N|^3)$. The complete algorithm of Lagrangean relaxation (denote as LGR algorithm) to solve (IP), which includes algorithms to solve each subproblem in Sec. III and LGR-Primal, has the same flowchart as in [6]. The computational complexity for LGR algorithm is $O(|S||N|^3)$ for each iteration.

V. COMPUTATIONAL EXPERIMENTS

The proposed algorithms for constructing a data aggregation tree are coded in C and run on a PC with PIV-2G. In LGR algorithm, *Max_Iteration_Number*[6] and *Improve_Threshold* [6] are set to 2000 and 30 respectively. The step size coefficient, δ [6], is initialized as 2 and is halved when the objective function value of the dual problem is not improved by iterations up to *Improve_Threshold*. The computational time for the following computational experiments is all within five minutes.

We assume that a sensor network operates in periodic mode (i.e. the sensor nodes periodically report information to the sink node). The network topology comprises N (= 150 in Fig. 3 and 4, up-to 250 in Fig. 5) sensor nodes randomly placed in a 1×1 square unit area. The cost of the energy consumption function, $e_n(r_n)$, is defined as the square of $100\times$Euclidean distance multiplied by energy consumption per millisecond when the sensor node is transmitting data. The set of all possible transmission radii of a sensor node n (R_n) is configured to begin from 0 to the *maximum communication radius* (e.g. 0.25 in Fig. 3). Elements in the radius set are increased by 0.01 successively. The CSMA related parameters (RTS, $SIFS$, θ) are the same setting as in [4]. To evaluate the solution quality of our proposed algorithm, we implement four existing algorithms for comparison. The SPT, GIT and CNS algorithms are proposed in [2] and the forth algorithm, CCA, is described in [5]. Note that all these four heuristics construct the data aggregation tree without considering the MAC layer collision effect. In addition, since CCA algorithm addresses the latency issue so that the data aggregation tree is a balance tree. From the computational results, the balance tree suffers from severe retransmission energy loss so that the total energy consumption is five times worse than our LGR algorithm. Therefore, we did not plot the results from CCA in Fig. 3-5 for better demonstrating the comparison between LGR algorithm and other three heuristics. Each plotted point in Fig. 3-5 is a mean value over 10 simulation results.

Fig. 3 shows the total energy consumption under different number of source nodes. It is shown that LGR algorithm can get best solution quality in terms of total energy consumption as compared with other heuristics. In addition, we observe that when the number of source nodes is large (e.g. 80, 90, 100), the solution quality LGR algorithm over other heuristics are even more significant. When the number of source nodes is large, we have a larger aggregation tree as compared to aggregation tree constructed by small number of data source nodes. Larger aggregation tree results in the greater probability for collision. Hence, the heuristics algorithms that do not address the MAC collision effect suffer from severe retransmission.

In Fig. 4, we examine the effect of communication radius on energy consumption with 90 source nodes. It is shown that LGR algorithm can still get best solution quality in terms of total energy consumption as compared with other heuristics. Interestingly, it is shown that large maximum communication radius did not give any advantage for MAC aware energy efficient data aggregation tree. On the contrary, large communication radius leads to severe collision that could jeopardize the advantages from data aggregation. Hence, with considering the tradeoff between collision and data aggregation, small communication radius should be the best communication radius setting for sensor node.

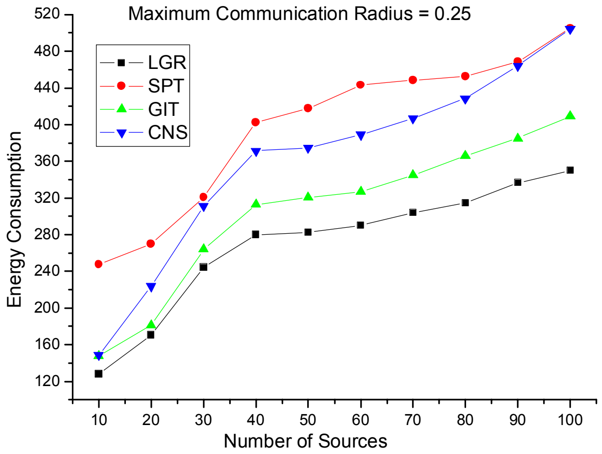

Figure 3. Energy consumption with different no. of source nodes

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.

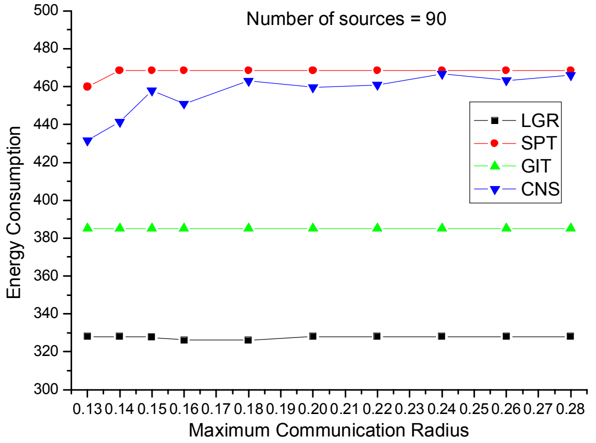

Figure 4. Energy consumption under different maximum radii

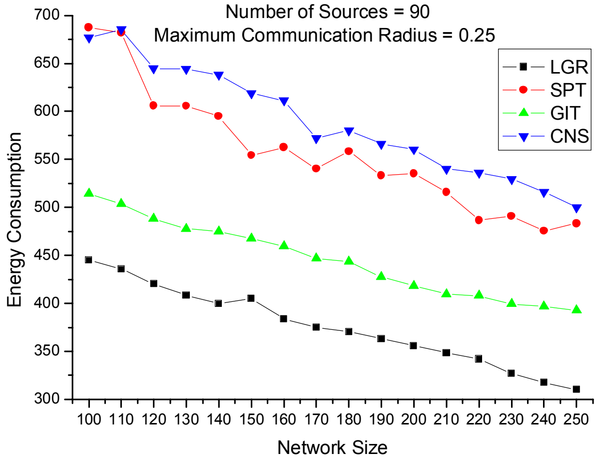

Figure 5. Energy consumption under different network sizes

Fig. 5 depicts the experiments evaluating the solution quality under different network sizes. LGR algorithm outperform than the other heuristics for all different network sizes. In addition, it is observed that total energy consumption is a monotonically decreasing function with respect to network size. When in large network size, we have a high density of sensor nodes in the fixed deployment area. In other words, sensor node could reach neighbor node with shorter transmission radius. Hence, we could conclude that shorter transmission radius is more energy efficient in MAC-DCR problem, which is the same insight as in Fig. 4,

We define the improvement ratio as (other approach −LGR)/(LGR)×100% to show the solution quality. From Table I, the improvement ratio of LGR over SPT, GIT and CNS is up to 93%, 27% and 63% respectively. Note that the improvement ratio of LGR over CCA is at least 500%.

TABLE I. Improvement Ratio

Improvement Ratio	Fig. 3	Fig. 4	Fig. 5
SPT	93	44	57
GIT	17	18	27
CNS	44	42	63

VI. CONCLUSION

Energy-efficient data-centric routing in WSN should carefully consider the retransmission energy loss due to MAC collision. In this paper, for the first time, we propose a rigorous nonlinear mathematical formulation for MAC aware energy efficient data centric routing problem in WSN where the objective function is to minimize the total energy consumption subject to data aggregation tree, routing assignment, transmission radius and data retransmissions constraints. The proposed solution approach is based on Lagrangean relaxation to construct a MAC aware energy-efficient data aggregation tree that jointly considers the tradeoff between data aggregation and data retransmission. According to the computational experiments, the proposed LGR algorithm outperforms other heuristics under all tested case. The other interesting observation is that small communication radius is better than large communication radius in terms of total energy consumption in MAC-DCR problem.

Acknowledgement

This work is supported (in part) by the National Science Council, Taiwan, under grant number NSC 94-2213-E-002-063 and 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] B. Krishnamachari, D. Estrin, and S.Wicker, "Modeling Data-Centric Routing in Wireless Sensor Networks.", *USC Computer Engineering Technical Report CENG* 02-14, 2002.

[3] 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.

[4] S.T. Sheu, T.-H Tsai and JH Chen, “MR 2 RP : The Multi-Rate and Multi-Range Routing Protocol for IEEE 802.11 Wireless Ad Hoc Networks”, *ACM/Kluwer Wireless Networks*, Vol. 9, No. 3, pp. 165-177, May 2003.

[5] S. Upadhyayula, V. Annamalai, and S. K. S. Gupta,” A Low-Latency and Energy-Efficient Algorithm for Convergecast in Wireless Sensor Networks”, *Proc. of IEEE GLOBECOM*, 2003.

[6] H. H. Yen, F. Y. S. Lin and S. P. Lin, “Efficient Data-Centric Routing in Wireless Sensor Networks”, *Proc. of IEEE ICC,* Vol. 5, pp. 3025-3029, 2005.

[7] W. Ye, J. Heidemann, and D. Estrin, “An Energy-Efficient MAC Protocol for Wireless Sensor Networks”, *USC/ISI Technical REPORT ISI-TR-543*, Sep. 2001.

[8] W. Ye and J. Heidemann, ”Medium Access Control in Wireless Sensor Networks”, *USC/ISI Technical Report ISI-TR-580*, Oct. 2003.

[9] X. H. Lin, Y. K. Kwok, and Vincent K. N. Lau, ”A New Power Control Approach for IEEE 802.11 Ad Hoc Networks”, *Proc. of the 14th IEEE PIMRC,* Vol. 2, pp. 1761-1765, Sep. 2003.