Multi-sink Data Aggregation Routing and Scheduling with Dynamic Radii in WSNs

Frank Yeong-Sung Lin and Yean-Fu Wen, *Student Member, IEEE*

Abstract— In wireless sensor networks (WSNs) with multiple sink nodes, energy is wasted on idle listening, redundant transmissions, and unnecessary power consumption. The total energy consumption may be minimized by properly scheduling communications in conjunction with aggregating data and dynamically adjusting radii. We propose near-optimal data aggregation routing and duty cycle scheduling heuristics, denoted by MDAR and O-MAC, which achieve energy efficiency and bound latency within a reasonable range. These heuristics outperform other general data aggregation routing heuristics (e.g., CNS, GIT, and SPT) and scheduling protocols (e.g., S-MAC and T-MAC) by 7%-45%, according to our experimental results.

Index Terms— Data aggregation, dynamic radii, energy efficiency, routing, scheduling, sensor networks.

I. INTRODUCTION

THE network lifetime of a WSN is dependent on battery capacity and energy efficiency. In this paper, we seek to prolong that lifetime from the physical layer up to the application layer, with focuses on (i) multiple sinks (outgoing information gateways), (ii) data aggregation routing, (iii) duty cycle scheduling and (iv) dynamic radii.

The assignment of data aggregation routing as a kind of reverse-multicast tree, which is also a Steiner tree, has been proven to be NP-complete [2] [4] [5]. Krishnamachari *et al.* [2] devised three routing heuristics, Shortest Path Tree (SPT), Center at the Nearest Source (CNS), and Greedy Incremental Tree (GIT), which solve the problem sub-optimally. For a network with multiple sinks, all sink nodes are connected to a pseudo destination by pseudo links, which reduces it to a single reverse-multicast tree, as shown in Fig. 1.

Sensor nodes periodically report information to one of the sink nodes. Energy is conserved by duty cycling, thereby reducing idle listening, the most energy wasteful process in the MAC protocol. Lu *et al.* [4] determined that neither the fixed duty cycle mechanism of S-MAC [10] nor the variable duty cycle by timeout of the more advanced T-MAC [9] are optimally designed. To eliminate such inherent sleep latency problems, we considered both near-optimal duty cycle scheduling and data aggregation to generate a new mechanism, O-MAC. An example of the duty cycle of nodes controlled in concert with the aggregation of data is shown in Fig. 1.

Energy consumption, $r^\alpha + c$, increases exponentially faster than the range of a node r, where α is signal attenuation, a constant between 2 and 4 [3]. Minimizing consumption
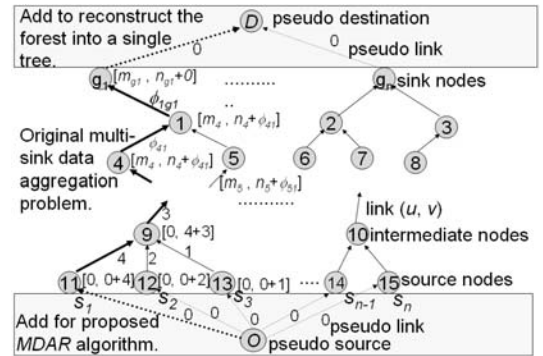
Fig. 1. An example of duty cycle scheduling for a data aggregation tree, link time value ϕ_{uv} signifying successful transmission to next node, and $[n_u, m_u + \phi_{uv}]$ denoting the times of earliest wake-up and aggregation completion (when successful transmission is made).

is accomplished by maintaining a dynamic radius; with an estimate of necessary power consumption, a node reduces its power consumption accordingly [6].

The remainder of the paper is organized as follows. The following section presents a mixed integer- and nonlinear- programming approach to the data aggregation routing problem, including the assignment of duty schedules and radii. Section III discusses heuristics for getting good feasible solutions to the problems. In Section IV, computational results are reported. Finally, in Section V, the conclusions are presented.

II. PROBLEM FORMULATION

Based on the above discussion, the problem is as follows:

Given: The given set of variables, shown in Table I.

Objective: Minimize the total energy consumed by a target transmission delivered to one of the sink nodes.

Subject to:

· One path from a source node to the pseudo destination D;

· Restrictions on the structure of trees in the form of three link constraints, i.e., (i) the number of out-degree links from source nodes; (ii) the number of out-degree links from relay nodes; and (iii) the summation of in-degree links to a selected sink node;

· The time for node-to-node communication based on the number of active neighbors and the CSMA/CA protocol; and

· Duty cycle scheduling.

To determine:

· The sink node to which a source node will route;

· A routing path from a source node to that sink node;

· The power range of each node;

· The time at which aggregation of sub-tree data is completed;

· The earliest time at which a node wakes up and begins aggregating data; and

· The time needed for a successful node-to-node transmission.

Accordingly, a WSN is modeled as a graph of connected nodes, $\Gamma(V, L)$. V represents the nodes distributed on a two-dimensional plane. L denotes the links that enable node v to

TABLE I
SUMMARY OF GIVEN PARAMETERS.

Notation	Description
V	The set of sensor nodes.
L	The set of links.
S	The set of source nodes, where $S \in V$.
G	The set of sink nodes.
P_{sD}	The set of candidate paths from source s to pseudo destination D.
M_i	An arbitrary large real number, where $i \in 1, 2, 3$.
λ	The occurrence rates of a target event.
θ_{uv}	The propagation time on link (u,v) to send a packet.
$\overline{B}$	The average random back-off time [1].
$\overline{N}$	The average NAV (Network Allocation Vector) time [1].
RTS	RTS transmission time [1].
CTS	CTS transmission time [1].
E_r	The energy consumption rate of a receiving or idle node [7].
$\delta_{p(uv)}$	The indicator function: 1 if link (u,v) is on path p; otherwise 0.
d_{uv}	The distance between node u and node v.
ϵ	Estimation error value (used for a guard time in synchronization).

TABLE II
SUMMARY OF DECISION VARIABLES.

Notation	Description
m_u	The longest aggregation time, measured from the source nodes of the aggregation sub-tree to node u.
n_u	The earliest time at which node u begins aggregating data.
y_{uv}	Equal to 1 if link (u,v) is in the tree; otherwise, equal to 0.
x_p	Equal to 1 if path p is selected to transmit a packet to the sink node; otherwise, equal to 0.
l_{uv}	The time necessary to transmit a packet from node u to node v.
ϕ_{uv}	The time necessary for a successful transmission from node u to node v when link (u,v) is used for the tree; otherwise, equal to 0.
z_{uv}	Equal to 1 if node v is within the transmission range of selected relay node u; otherwise, equal to 0.
r_u	The transmission range of node u, where $r_u \in R_u$.
$e_u(r_u)$	The energy consumption function for node u, a function of the transmission radius of u.

receive a signal from node u. Tables I and II list the given parameters and the decision variables, respectively.

The total energy consumption of a network forms the basis of the objective function (IP), including all energy consumed when listening/receiving (the first term of the IP) and when transmitting (the second term of the IP). Minimizing the objective function implies maximal network lifetime.

$$Z_{IP} = min \sum_{u \in V} [(m_u - n_u)E_r + e_u(r_u) \sum_{v \in V} \phi_{uv}], \quad (IP)$$

which is subject to:

A. Path constraints

One path must be found to connect a source node with the pseudo destination, D. This constraint requires:

$$\sum_{p \in P_{sD}} x_p = 1, \forall s \in S \quad (1)$$

Once a path p is selected, if a link (u,v) is on the path, then the decision variable y_{uv} is set to 1, as described by (2):

$$\sum_{p \in P_{sD}} x_p \delta_{p(uv)} \leq y_{uv}, \forall s \in S; u, v \in V \quad (2)$$

B. Link constraints

When the problem is reduced to determination of a single reverse-multicast tree, roots given by the pseudo destination D, three link constraints restrict the structure of such a tree, ensuring that all sensor data is delivered and by a single path.

1) A single out-going link for source node s:

$$\sum_{v \in V} y_{sv} = 1, \forall s \in S. \quad (3)$$

2) No more than one out-going link for each node:

$$\sum_{v \in V} y_{uv} \leq 1, \forall u \in V. \quad (4)$$

3) At least one in-going link to D:

$$\sum_{u \in V} y_{uD} \geq 1. \quad (5)$$

C. Node-to-node communication time constraints

Equation (6), which refers to [8], denotes the time needed to transmit from node u to node v by the CSMA//CA protocol [1]. Note no DIFS and SIFS spacing for WSNs.

$$l_{uv} = \frac{(e^{-\lambda \sum_{j \in V} z_{ju}} (RTS + CTS + \bar{B}) + \bar{N})}{e^{-\lambda \sum_{j \in V} z_{ju}} e^{-(RTS+2\theta) \sum_{j \in V} z_{jv}}} - \bar{N}, \forall u, v \in V \quad (6)$$

where decision variable z_{ju} is 1 when link (j,u) is selected (i.e. $y_{ju} = 1$), as shown in (7), where M_1 is equal to the maximum radius. However, z_{ju} must be equal to 0 when no data is to be transmitted by node u (i.e. $r_u = 0$), as in (8). The details of these two equations are described as the following.

$$\frac{r_u - d_{ju}}{M_1} + (y_{ju} - 1) \leq z_{ju}, \forall j, u \in V. \quad (7)$$

$$z_{ju} d_{ju} \leq r_u, \forall j, u \in V. \quad (8)$$

Constraint (9) determines whether the time necessary for node-to-node transmission ϕ_{uv} is set or fixed to zero. The value of ϕ_{uv} is calculated when a link (u,v) is selected, except in the case of a pseudo link; if a link is not selected, it has no weight. Note that M_2 is upper bound on l_{uv}.

$$l_{uv} - M_2(1 - y_{uv}) \leq \phi_{uv}, \forall u, v \in V, v \neq D \quad (9)$$

D. Duty cycle scheduling (O-MAC) constraints

Equation (10) limits the time at which all incoming flow from nodes must be received by aggregating nodes.

$$m_v + \phi_{vu} + \varepsilon \leq m_u, \forall u, v \in V \quad (10)$$

A node involved in an aggregation tree is subject to (11). Wake up must occur prior to the aggregation of data. Note that M_3 is an upper bound on n_u.

$$n_u \leq m_v + M_3(1 - y_{vu}), \forall u, v \in V \quad (11)$$

III. PROPOSED ALGORITHM

The proposed heuristic, MDAR (Multi-sink Data Aggregation Routing), includes duty cycles scheduled by O-MAC procedures, shown as the following. For an example of MDAR decision-making, refer back to Fig. 1. The first iteration finds the path from O to D, O-11- 9-$\cdots$-4-1-$g1$-D. Nodes on the path belong to set T. Pseudo links $(11,D)$, $(9,D)$,$\cdots$, $(4,D)$ and $(1,D)$ with weight 0 are added. The trees are marked iteratively until all source nodes in the network have been linked. The duty cycle $[n_u, m_u + \phi_{uv}]$ for O-MAC and power range r_u
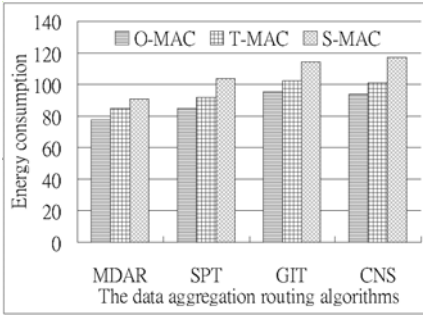
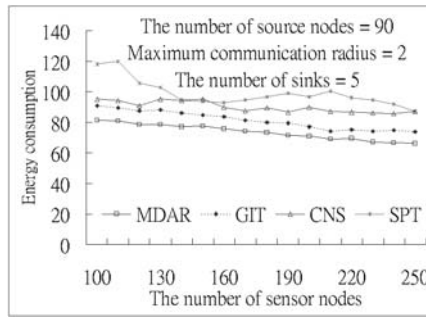Fig. 2.   Energy consumed by data aggregation rout-ing heuristics combined with scheduling heuristics.

Fig. 3.   Effect of network size on energy consump-tion for heuristics.
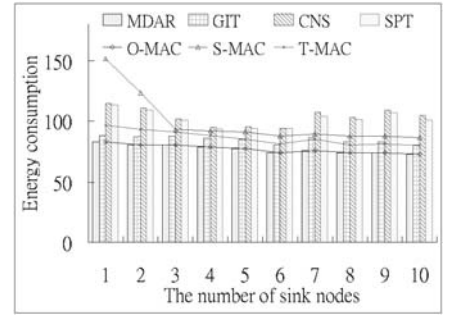
Fig. 4.   Effect of different numbers of sink nodes on energy consumption.

of each node u are calculated for each sensor node. The time complexity for these calculations is $O(|S||V|^2)$.

Step 1 Initially, let links be weighted by energy consumption rate $w_{uv} = d_{uv}^\alpha + c$; pseudo links connect the source nodes to the pseudo source, initial weights 0.

Step 2 Dijkstra's algorithm is used to find the shortest path from the pseudo source to the pseudo destination.

Step 3 Once that path has been determined, the nodes on the path are marked. Pseudo links, weight 0, are inserted between marked nodes and the pseudo destination. The weight of pseudo links from the pseudo source are set to infinity.

Step 4 Steps 2-3 are repeated until all source nodes are marked.

Step 5 All pseudo links and pseudo nodes are removed, and at least one reverse-multicast tree has been generated. Once the set $\{x_p\}$ has been determined, the set of links $\{y_{uv}\}$ is determined. The link delays ϕ_{uv} and the power ranges r_u are calculated based on $\{y_{uv}\}$.

Step 6 The earliest wake up time $\{n_u\}$ and latest aggregation time $\{m_u\}$, are calculated along sequentially visited nodes by the DFS (Depth First Search) algorithm. Once $\{m_u\}$ and $\{n_u\}$ are determined, the duty cycle schedule $[n_u, m_u + \phi_{uv}]$ is determined for each node u. The total energy consumption per cycle is then calculated accordingly.

IV. EVALUATION AND EXPERIMENTAL RESULTS

For evaluation purposes, networks in this experiment were comprised of N sensor nodes with maximum radius of 2.5 units randomly distributed over an area 10 by 10 units square. The number of source nodes was 90, the number of sensor nodes was 150, and the number of sink nodes was 5 for Figs. 2 and 4. CSMA related parameters (e.g., RTS, CTS, and etc.) were set as in [8]. The program was written in C++ and run in Linux 5.2.1.

Fig. 2 shows the results for combinations of data scheduling and data aggregation routing heuristics. By scheduling wake-up duty cycles that suit the nature of the aggregation tree, O-MAC outperforms S-MAC and T-MAC by at least 6.8%. S-MAC permits an inordinate amount of time for idle listening. With the aid of a timeout mechanism, T-MAC outperforms S-MAC; however, more time is still spent on idle listening than in O-MAC. Fig. 3 shows that in networks running O-MAC, MDAR consistently outperforms the other routing heuristics by 7.5%-32.6%.

The duration of a duty cycle is determined by a tradeoff between data aggregation and path length, with closer sink nodes allowing for shorter duty cycles. Thus, there exists an inverse relationship: total energy consumption decreases as

the number of sink nodes increases. The experimental results shown in Fig. 4 demonstrate that the network performance achieved by other data aggregation routing and scheduling heuristics may be surpassed by 8.7%-45.4% by using both MDAR and O-MAC. Note that for routing performance, O-MAC is used for MAC; for scheduling performance, MDAR for the routing algorithm.

V. CONCLUSION

The energy efficiency of WSNs can be improved by data aggregation routing, the reduction of idle listening, and dynamic control of radius. We proposed an mixed integer- and nonlinear-mathematical formulation that addresses these factors. The experimental results show, for multi-sink WSNs, our proposed O-MAC and MDAR heuristics to significantly outperform other heuristics, even in cases where O-MAC operates in combination with a standard aggregation routing heuristic, or where MDAR is combined with a standard scheduling protocol.

REFERENCES

[1] ANSI/IEEE, "802.11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications," IEEE Standard 802.11, 2000.

[2] B. Krishnamachari, D. Estrin, and S. Wicker, "Modelling data-centric routing in wireless sensor networks," USC Computer Engineering Technical Report CENG 02-14, 2002.

[3] J. Cartigny, D. Simplot, and I. Stojmenovic, "Localized minimum-energy broadcasting in ad-hoc networks," in *Proc. IEEE INFOCOM 2003*, vol. 3, pp. 2210-2217.

[4] G. Lu, N. Sadagopan, B. Krishnamachari, and A. Goel, "Delay efficient sleep scheduling in wireless sensor network," in *Proc. IEEE INFOCOM 2004*, vol. 4, pp. 2470-2481.

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

[6] H. H. Yen, F. Y. S. Lin and S. P. Lin, "Energy-efficient data-centric routing in wireless sensor networks," *IEICE Trans. Commun.*, vol. E88-B, pp. 4470-4480, Dec. 2005.

[7] M. Stemm and R. H. Katz, "Measuring and reducing energy consumption of network interfaces in hand-held devices," *IEICE Trans. Commun.*, vol. E80-B, pp. 1125-1131, Aug. 1997.

[8] S. T. Sheu, T. H. Tsai, and J. H. Chen, "MR^2RP: the multi-rate and multi-range routing protocol for IEEE 802.11 wireless ad hoc networks," *ACM/Kluwer Wireless Networks*, vol. 9, pp. 165-177, May 2003.

[9] T. V. Dam and K. Langendoen, "An adaptive energy-efficient MAC protocol for wireless sensor networks," in *Proc. ACM SenSys 2003*, pp. 171-180.

[10] W. Ye, J. Heidemann, and D. Estrin, "An energy-efficient MAC protocol for wireless sensor networks," in *Proc. IEEE INFOCOM 2002*, vol. 3, pp. 1567-1576.