Distributed Minimum Energy Data Gathering and Aggregation in Sensor Networks

Kevin Yuen, Baochun Li, Ben Liang
Department of Electrical and Computer Engineering
University of Toronto
10 King's College Road
Toronto, Ontario, M5S 3G4, Canada
{*yuenke, bli*}*@eecg.toronto.edu, liang@comm.utoronto.ca*

Abstract— **In this paper, we propose an effective distributed algorithm to solve the minimum energy data gathering (MEDG) problem in wireless sensor networks. The problem objective is to find an optimal transmission structure on the network graph, such that the total energy consumed by the sensor nodes is minimized. We formulate the problem as a non-linear optimization problem. The formulation considers in-network data aggregation and respects the capacity of the wireless shared-medium. We apply Lagrangian dualization technique on this formulation to obtain a subgradient algorithm for computing the optimal transmission structure. The subgradient algorithm is asynchronous and amenable to fully distributed implementations, which corresponds to the decentralized nature of sensor networks.**

I. INTRODUCTION

Recent technological advances have enabled the production of low-cost sensor nodes. These sensor nodes are small in size, and are equipped with limited sensing, processing, and transmission capabilities. They can be deployed in large numbers to form sensor networks with the ability of distributed wireless sensing. In practice, sensor nodes are deployed in an ad-hoc fashion over the area of interest. After the deployment, sensor nodes collect data from their surroundings, encode the data, and transmit them to the sink node via wireless channels. In addition, intermediate sensor nodes can be used as relays. The sink node is a specialized node in the network responsible for receiving data, and serves as a gateway between the sensor network and the wired backbone network.

Many applications for sensor networks, such as target tracking [1] and habitat monitoring [2], involve monitoring a remote or hostile field. Sensor nodes are assumed to be inaccessible after deployment for such applications and thus their batteries are irreplaceable. Also, due to the small size of sensor nodes, they carry limited battery power. Thus, energy is a scarce resource that must be conserved to the extent possible in sensor networks.

In this context, the objective of the MEDG problem is to find the optimal transmission structure on the network graph, such that the total energy consumed in transporting the data from the sensor nodes to the sink node is minimized. If the data collected by the sensor nodes are independent and the wireless links have unlimited bandwidth capacities, then the solution to the MEDG problem is trivial – each sensor node can transmit its collected data to the sink node via the minimum energy path.

However, in sensor networks, sensor nodes are often densely deployed. Nearby sensor nodes have overlapping sensing ranges and the data they collected are either redundant or correlated. This data correlation can be exploited to reduce the amount of data transmitted in the network with data aggregation techniques, resulting in energy savings. Data aggregation is introduced by Estrin *et al.* [3] as an essential paradigm for wireless routing in sensor networks. The concept is to aggregate the data originating from different sensor nodes during their transmission to the sink node. In this paper, we assume a perfect data correlation model, where intermediate sensor nodes perform simple aggregation functions to aggregate multiple input packets into a single output packet. To achieve minimum energy data gathering, the optimal transmission structure should allow maximum aggregation of data.

Besides the correlated sensor nodes, the optimal transmission structure also depends on the link capacities. In any practical networks, there are capacity limitations on the links and interference among competing signals. In wireline networks, there is time-dependent contention, where two signals compete with each other if they both arrived at the router at the same time. The effect of interference in wireline networks is well studied, but they are not applicable in the context of sensor networks. As a variation of wireless ad hoc networks, sensor networks have the unique characteristic of location-dependent contention, in addition to time-dependent contention. Signals will compete with each other if nearby sensor nodes access the wireless shared-medium at the same time. In this paper, we model the location-dependent contention as constraints according to the protocol model of packet transmission in wireless networks [4].

In addition to data correlation and channel interference, there are other factors that should be considered when solving the MEDG problem:

- *Distributed solutions:* With centralized solutions, participating nodes need to transmit detailed status information repeatedly across the network to a central computation node. Although centralized approaches can achieve re-

sults closer to the optimal, they are generally not feasible.

– *Asynchronous network environments:* Due to the ad-hoc infrastructure of sensor networks, it is expensive in terms of communication overhead to synchronize the nodes.

In this paper, we propose an efficient algorithm to solve the MEDG problem. To obtain the optimal solution, it is a natural decision to employ optimization techniques. We formulate the MEDG problem as a non-linear optimization problem. This formulation considers in-network data aggregation and accounts for the effect of location-dependent contention. As a result, our solution is guaranteed to be supported by the wireless shared-medium. After applying Lagrangian dualization on the formulation, we utilize a price-based resource allocation strategy to find the optimal transmission structure. Price signals are used in the algorithm to reflect the congestion status of the network. To facilitate its deployment in practical sensor networks, our algorithm is asynchronous and amenable to distributed implementations.

Data gathering with correlated sources in sensor networks and resource allocation with capacity limitations in wireless networks have been separately studied in previous literature. The main contribution of this paper is to propose a distributed solution to the MEDG problem that considers both topics simultaneously. Our algorithm finds a minimum energy transmission structure that allows maximum data aggregation and respects the wireless link capacities. To the best of our knowledge, no previous works have addressed the MEDG problem with all of the factors above.

The remainder of the paper is organized as follows. In Section II, we present the optimization formulation for the MEDG problem. In Section III, we construct an efficient distributed algorithm to solve the formulation. Numerical results from simulations are presented in Section IV. Finally, we discuss related work in Section V, and conclude the paper in Section VI.

II. PROBLEM FORMULATION

A. Network Model

The wireless sensor network is modeled as a graph $G = (V, E)$, where V is the set of nodes and E is the set of bidirectional wireless links. Let S_N denotes the set of sensor nodes and S_K denotes the sink node. Then, $V = S_N \cup S_K$. All sensor nodes have a maximum transmission range of r_{tx}. Let d_{ij} denote the distance between node i and node j. A bidirectional link $(i, j) \in E$ exists if $d_{ij} \leq r_{tx}$. Each link is associated with an energy cost e_{ij}, referring to the energy consumed per unit flow on link (i, j). Moreover, f_{ij} represents the flow rate of link (i, j). Here, the flow vector $[f_{ij}]_{\forall (i,j) \in E}$ are the variables that can be adjusted in order to minimize the optimization objective stated in the next subsection.

There are various models for sensor networks. In this work, we focus on a sensor network environment where:

- Each sensor node has a continuous data delivery model. The sensor nodes periodically sense their surroundings and always have data to transmit in each round of communication.

- Sensor nodes have the ability to adjust their transmission power depending on the distance of transmission.
- Sensor nodes and the sink node are stationary.
- Simple aggregation functions are utilized by the sensor nodes to aggregate the data they received with their own data. Consequently, each sensor node produces only one packet regardless of how many packets it has received in each round of communication. This perfect data correlation model is also utilized in [5], [6], [7].
- Each node in the network has knowledge of its own location.

B. Optimization Objective

Given a transmission structure, the flow vector $[f_{ij}]_{\forall (i,j) \in E}$ can be determined, and the total energy consumed by each link equals $e_{ij} \cdot f_{ij}$. The objective of the MEDG problem is to minimize the total energy consumed in the network:

$$\textbf{Minimize} \quad \sum_{(i,j) \in E} e_{ij} \cdot f_{ij}. \tag{1}$$

C. Constructing the Minimum Spanning Tree

With our network model, it is shown in [3] that the optimal number of transmissions required per round of communication equals to the number of edges in the minimum spanning tree (MST) on the network graph. Thus, for any graph G, its optimal transmission structure is equivalent to its MST. For any proper subset $S \subset V$, let $\delta(S) = \{(i, j) : \text{exactly one endpoint of } (i, j) \text{ is in } S\}$. Then it is possible to construct the MST of G by combining the optimization objective with the following constraints:

$$\sum_{(i,j) \in \delta(S)} f_{ij} \geq 1, \quad \forall S \subset V, \tag{2}$$

$$f_{ij} = \{0, 1\}, \quad \forall (i, j) \in E. \tag{3}$$

Note that a simple MST solution cannot solve the MEDG problem because the MST may not be supported by the wireless shared-medium.

D. Channel Contention Constraints

In any wireless ad hoc networks, packet transmission is subject to location-dependent contention, where nearby data flows compete for the capacity of the wireless shared-medium. In the literature, there exist two models for packet transmission in wireless networks [4]. They are generally referred to as the *protocol model* and the *physical model*. According to the protocol model, all links adjacent to node k will interfere with bidirectional link (i, j) if $d_{ki} < (1 + \triangle)d_{ij}$ or $d_{kj} < (1 + \triangle)d_{ij}$, where the quantity $\triangle > 0$ specifies a guard zone. In this paper, we utilize the protocol model of packet transmission, and derive Ψ_{ij} for each link $(i, j) \in E$ as the cluster of links that cannot transmit simultaneously with link (i, j). The notation of cluster is used here as a basic resource unit, as compared to individual links in the traditional wireline networks. In wireline networks, data flows compete for the capacity of individual links. However, in the case of

sensor networks, the capacity of a wireless link is interrelated with other wireless links in its cluster. Consequently, data flows compete for the capacity of individual clusters, which is equivalent to the capacity of the wireless shared-medium. A flow vector $[f_{ij}]_{\forall (i,j) \in E}$ is supported by the shared-medium if the channel contention constraints below hold.

$$f_{ij} + \sum_{(p,q) \in \Psi_{ij}} f_{pq} \leq C, \quad \forall (i,j) \in E, \qquad (4)$$

where C is defined as the maximum rate supported by the wireless shared-medium. We note that the channel contention constraints are generic, since they can accommodate other models of packet transmission instead of the protocol model.

E. Non-Linear Programming Formulation

Combining the optimization objective with the introduced constraints, the MEDG problem can be modeled as an optimization problem as follows:

$$\textbf{Minimize} \quad \sum_{(i,j) \in E} e_{ij} \cdot f_{ij} \quad \textbf{s.t.} \qquad (5)$$

$$\sum_{(i,j) \in \delta(S)} f_{ij} \geq 1, \quad \forall S \subset V, \qquad (6)$$

$$f_{ij} + \sum_{(p,q) \in \Psi_{ij}} f_{pq} \leq C, \quad \forall (i,j) \in E, \qquad (7)$$

$$f_{ij} = \{0,1\}, \quad \forall (i,j) \in E. \qquad (8)$$

The number of constraints required to construct the MST grows at an exponential rate in relation to the number of nodes. This makes the optimization problem described above a non-linear programming formulation.

III. DISTRIBUTED SOLUTION

Many algorithms have been proposed to solve optimization problems, such as simplex, ellipsoid and interior point methods. These algorithms are efficient in the sense that they can solve large instance of optimization problems in a few seconds. However, they have the disadvantage of being inherently centralized, which implies that they are not applicable for distributed deployment. In this section, we propose our distributed solution to the MEDG formulation.

A. Lagrangian Dualization

The MEDG formulation resembles a resource allocation problem, where the objective is to allocate the limited bandwidth capacities of the clusters to the data flows originating from the sensor nodes. Previous research works in wireline networks [8], [9] have shown that a price-based strategy is an efficient means to arbitrate resource allocation. In this strategy, each link is treated as a basic resource unit. A shadow price is associated with each link to reflect the relation between the traffic load of the link and its capacity. Based on the notation of maximal cliques, Xue *et al.* [10] extend this price-based resource allocation framework to respect the unique characteristic of location-dependent contention in wireless

networks. Due to the complexities in constructing maximal cliques, the notation of cluster as defined in Section II is used as the basic resource unit in this paper. Each cluster is associated with a shadow price, and the transmission structure is determined in response to the price signals, such that the aggregated price paid by the data flows is minimized. It is revealed from previous research that at equilibrium, such price-based strategy may achieve global optimum, leading to the optimal utilization of resource.

To solve the MEDG formulation with a price-based strategy, we relax the channel contention constraints (7) with Lagrangian dualization to obtain the Lagrangian dual problem:

$$\textbf{Maximize} \quad \textbf{LS}(\beta) \quad \textbf{s.t.} \quad \beta \geq 0. \qquad (9)$$

By associating price signals or Lagrangian multipliers β_{ij} with the channel contention constraints, the Lagrangian dual problem is evaluated via the Lagrangian subproblem:

$$\textbf{Minimize} \quad \sum_{(i,j) \in E} e_{ij} \cdot f_{ij} + \beta_{ij} \cdot (f_{ij} + \sum_{(p,q) \in \Psi_{ij}} f_{pq} - C) \quad \textbf{s.t.}$$
$$(10)$$

$$\sum_{(i,j) \in \delta(S)} f_{ij} \geq 1, \quad \forall S \subset V, \qquad (11)$$

$$f_{ij} = \{0,1\}, \quad \forall (i,j) \in E. \qquad (12)$$

We further define Φ_{ij} as the set of clusters that link (i,j) belongs to. Recall Ψ_{pq} is the cluster of links that cannot transmit when link (p,q) is active. For any link (i,j) that interferes with link (p,q), link (i,j) belongs to the cluster of link (p,q). Thus, for any links (i,j) and (p,q), $(p,q) \in \Phi_{ij}$ iff $(i,j) \in \Psi_{pq}$. The Lagrangian subproblem can be remodelled using this notation:

$$\textbf{Minimize} \quad \sum_{(i,j) \in E} f_{ij}(e_{ij} + \beta_{ij} + \sum_{(p,q) \in \Phi_{ij}} \beta_{pq}) - \beta_{ij}C \quad \textbf{s.t.}$$
$$(13)$$

$$\sum_{(i,j) \in \delta(S)} f_{ij} \geq 1, \quad \forall S \subset V, \qquad (14)$$

$$f_{ij} = \{0,1\}, \quad \forall (i,j) \in E. \qquad (15)$$

The objective function of the remodelled Lagrangian subproblem specifies that the weight of each link is equal to the sum of its energy and capacity cost. And the capacity cost is equal to the Lagrangian multiplier of the link plus the sum of the Lagrangian multipliers in Φ_{ij}. This is intuitive since when link (i,j) is active, any links in the set Φ_{ij} cannot transmit due to interference. Hence the actual price to pay for accessing link (i,j) should equal to the total price for accessing link (i,j) and all links in Φ_{ij}.

Since the capacity constraints are relaxed, we observe that the solution of the Lagrangian subproblem requires each sensor node to transmit its aggregated data along its MST branch that leads to the sink node. As a result, finding the MST on the network graph can solve the Lagrangian subproblem. This can be achieved using a distributed MST algorithm [11].

B. Subgradient Algorithm

The subgradient algorithm is an efficient iterative algorithm to solve the Lagrangian dual problem. We start with a set of initial non-negative Lagrangian multipliers $\beta_{ij}[0]$. A possible choice for the initial Lagrangian multipliers can be zeroes, where we assume there is no congestion in the network. In this case, the initial transmission structure chosen by the algorithm will be the actual MST without any adjustments on the link weights. During each iteration k, given current Lagrangian multiplier values $\beta_{ij}[k]$, we solve the Lagrangian subproblem by finding the MST on the network graph, where the weight of a link equals to the sum of its energy cost, its Lagrangian multiplier, and the Lagrangian multipliers of the clusters that this link belongs to. Using the new flow vector $[f_{ij}[k]]_{\forall(i,j)\in E}$ obtained from the Lagrangian subproblem, we update the Lagrangian multipliers by:

$$\beta_{ij}[k+1] = \max(0, \beta_{ij}[k] + \theta[k](f_{ij}[k] + \sum_{(p,q)\in\Psi_{ij}} f_{pq}[k] - C)),$$

where θ is a prescribed sequence of step sizes. Note from the above equation that the Lagrangian multipliers varies depending on the value of $(f_{ij} + \sum_{(p,q)\in\Psi_{ij}} f_{pq} - C)$, which represents the amount of the capacity violation within a cluster. When the violation of a cluster is positive, there are data flows traveling in the cluster that are not supported by the wireless shared-medium. Conversely, when the violation of a cluster is negative, there exists free bandwidth in the cluster that is not utilized by the data flows.

The selection of step sizes plays an important role in the subgradient algorithm. If the step sizes are too small, then the algorithm has a slow convergence speed. If the step sizes are too large, then β_{ij} may oscillate around the optimal solution and the algorithm fails to converge. However, the convergence is guaranteed when θ satisfies the following conditions [12]:

$$\theta[k] \geq 0, \lim_{k\to\infty} \theta[k] = 0, \text{ and} \sum_{k=1}^{\infty} \theta[k] = \infty. \qquad (16)$$

In this paper, we use the following sequence of step sizes, $\theta[k] = \frac{a}{(b+ck)}$, where a, b, and c are positive constants.

C. Distributed Algorithm

Based on the subgradient algorithm, we design our distributed algorithm for the MEDG problem. Each bidirectional link (i,j) is delegated to its lower ordered node, assuming each node has a distinct identity. All computations related to link (i,j) will be executed on its lower ordered node. We now summarize our distributed algorithm:

1) Choose initial Lagrangian multiplier values $\beta_{ij}[0]$, for all links $(i,j) \in E$.
2) For the kth iteration, determine the weight of each link as $(e_{ij} + \beta_{ij}[k] + \sum_{(p,q)\in\Phi_{ij}} \beta_{pq}[k])$.
3) Compute the MST on the network graph using a distributed MST algorithm.
4) Based on the MST obtained, compute $f_{ij}[k+1]$, for all links $(i,j) \in E$.

5) Update Lagrangian multipliers $\beta_{ij}[k+1] = \max(0, \beta_{ij}[k] + \theta[k](f_{ij}[k] + \sum_{(p,q)\in\Psi_{ij}} f_{pq}[k] - C))$, where $\theta[k] = \frac{a}{(b+ck)}$, for all links $(i,j) \in E$.
6) For each link (i,j), send $\beta_{ij}[k+1]$ to all links in Ψ_{ij} and send $f_{ij}[k+1]$ to all clusters in Φ_{ij}.
7) Repeat steps 2 to 6 until convergence.

In asynchronous network environments, nodes with different computation speeds will execute the algorithm at varying paces. As a result, the links may not always have the most recent price and rate information of other links due to delayed or out-of-order updates. To accommodate asynchronous updates, we introduce the *partial asynchronism model* that will be used in the practical implementation of our algorithm. The partial asynchronism model makes the following assumption:

There exists a positive integer B such that:

- For every link (i,j), the time between consecutive updates is bounded by B for both price and rate updates.
- One-way communication delays between any two nodes is at most B time steps.

This partial asynchronism model is first discussed in [13]. Later, it is adapted by Low *et al.* [9] in wireline networks and Xue *et al.* [10] in wireless networks. In [10], a technique is proposed to improve the price-based strategy to accommodate asynchronous updates. At time t, instead of the most recent information, a link may only recognizes a sequence of recent updates. The idea of the technique is for the links to estimate the price and rate information by computing the average of the sequence from time $t - B$ to t. To improve the estimation, a moving average can be utilized with a heavier weight assigned to the more recent updates. From their simulations, it is shown that the price-based strategy will converge the fastest when the entire weight is assigned to the most recently received update. Moreover, they prove that with sufficiently small step sizes θ, the algorithm converges to the global optimum in asynchronous network environments. We adapt this technique in our implementation, allowing each link to estimate the price and rate information of other links based on the most recently received update.

IV. SIMULATION

A. Simulation Environments

We study the distributed MEDG algorithm in three different simulation environments. In our first environment, we solve the non-linear MEDG formulation presented in Section II directly with the non-linear solver LOQO 6.02 called from AMPL scripts. The computations are performed on the NEOS server for optimization [14]. For convenience, we refer to this environment as *optimal*. The purpose of the *optimal* environment is to provide us with an optimal solution to the MEDG problem, which is produced with a centralized scheme and global information. Since there are no other existing routing algorithms that consider data aggregation and wireless channel interference simultaneously in the literature, the optimal solution will serve as a reference in the performance evaluation of our distributed MEDG algorithm. The second

simulation environment is referred to as *synchronous*. In this environment, the MEDG problem is solved with a synchronous implementation of the algorithm. The local clocks on the nodes are synchronized, such that they will simultaneously execute the algorithm at every time step ($t = 1, 2, 3 \ldots$). Bounded communication delay is assumed where the price and rate updates will arrive at their destinations before the next time step. The third simulation environment, referred to as *asynchronous*, is based on the partial asynchronism model presented in Section III. The model assumes the existence of an integer B that bounds the time between consecutive price and rate updates. To implement this environment, a timer initially set to a random integer between 0 to B is maintained on each node. At each time step, the nodes decrease their timers by 1. When the timer reaches 0, the corresponding node executes an iteration of the algorithm, which includes sending the price and rate updates. After the execution, the node resets its timer to a random integer between 0 to B. In the *asynchronous* environment, update messages may be delayed or out-of-date.

Except for the *optimal* environment, the distributed MEDG algorithm is implemented with the C++ programming language. In all of our experiments, we assume that the transmission range is identical to the interference range. Furthermore, we follow a generic signal path loss model and let $e_{ij} = d_{ij}^2$.

B. Convergence Behaviour

In this subsection, we study the convergence behaviour of our algorithm. To this end, we generate a network topology with 100 nodes randomly distributed on a square field of 50m × 50m, and each node has a transmission range of 10m. In each time step, all sensor nodes will transmit one unit of data destined to the sink node. We set the capacity of the wireless shared-medium to 27 units, which correspond to the minimum capacity that a feasible solution exists. The algorithm is executed for 300 time steps under the *synchronous* and the *asynchronous* environments, and the total energy consumed at each time step is recorded. Fig. 1 shows that our algorithm converges to the identical optimal value in both environments. While the algorithm converges rapidly in the *synchronous* environment, the result from the *asynchronous* environment fluctuates for around 50 time steps before converging toward the optimal value.

C. Performance Evaluation

In our second study, we evaluate the performance of our algorithm by comparing the *asynchronous* environment against the *optimal* environment. We generate multiple network topologies with fewer nodes. The reason for moving to smaller topologies is due to the exponential growth of constraints in the non-linear MEDG formulation. The number of constraints grows at 2^n, where n is the number of nodes in the network. Due to the limitation of the solver, we are only able to simulate the *optimal* environment with at most 15 nodes. The nodes in the topologies are randomly distributed on a square field of 50m × 50m. The transmission range is
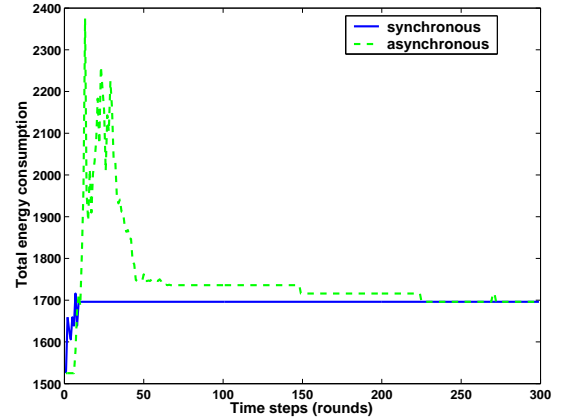

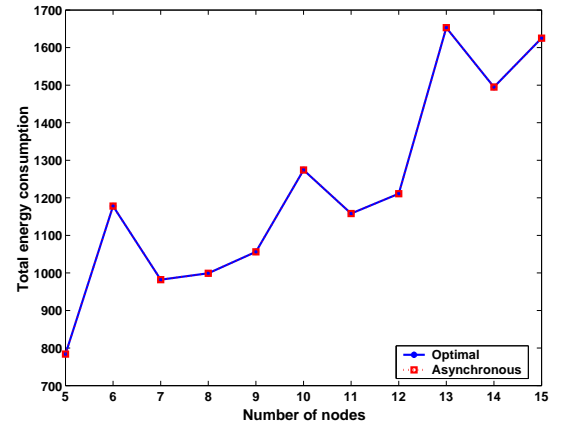Fig. 1. Convergence in asynchronous network environments.

Fig. 2. Optimal vs. asynchronous simulation environment.

30m and the capacity of the wireless shared-medium is set to the minimum value, such that a feasible solution exists. The total energy consumed by the different environments at convergence is recorded, for one instance of the random network topology given the number of nodes. We observe from Fig. 2 that our algorithm achieves the optimal solution for all topologies that were simulated. This result validates the claim that our distributed algorithm can find the minimum energy data gathering tree that is supported by the wireless shared-medium.

V. RELATED WORK

The problem of energy efficient routing in sensor networks has been investigated with optimization techniques in many previous works [15], [16], [17], [18]. Chang *et al.* [15] have formulated a flow-based linear programming problem to maximize the network lifetime. In [16], the optimization model minimizes energy consumption and considers the channel contention constraints associated with the wireless shared-medium. Krishnamachari *et al.* [17] propose another optimization formulation to maximize the raw data arrived at the sink node, subject to flow, fairness, energy, and capacity constraints. Boyd *et al.* [18] study the simultaneous routing

and power allocation problem for wireless data networks. In [17] and [18], the optimization formulations utilize the physical model [4] of packet transmission in wireless networks to model the channel contention constraints. However, the resulting constraints are non-convex, which makes the resulting formulations difficult to solve. In this paper, we model channel contention as linear constraints based on the protocol model. Furthermore, even though these works generally save energy, they do not consider the additional energy savings that can be achieved by exploiting data correlation among the sensor nodes.

The concept of data aggregation is to exploit data correlation among the sensor nodes by eliminating redundancy. Consequently, there are fewer transmissions in the network, which results in energy savings. In [6], Kalpakis *et al.* have formulated the maximum lifetime data gathering problem as an integer program with linear constraints, taking into consideration data aggregation, and present a polynomial-time algorithm to solve the problem. In [7], Goel *et al.* consider the joint treatment of data aggregation and transmission structure. The data aggregation function utilized by the intermediate nodes is modeled with concave, non-decreasing cost functions. Single-input coding is considered in [19], where intermediate nodes can aggregate their collected data with the side information provided by another node. In contrast, data aggregation with multi-input coding is only performed when all input information from multiple nodes is available. Multi-input coding is often employed by cluster-based routing algorithms such as LEACH [20] and PEGASIS [21]. In the model of LEACH, the cluster heads are responsible for aggregating all data arriving from the nodes in their respective cluster into a single packet. Instead of clusters, the PEGASIS algorithm finds chains of nodes, and the head node of each chain aggregates data from other nodes in the chain. Although these works yield satisfactory performance, they do not incorporate the effect of wireless channel interference in their design.

VI. Conclusion

In this paper, we have presented an efficient distributed algorithm for the MEDG problem in sensor networks. We design an optimization formulation for the MEDG problem, and show that the optimal transmission structure depends on the data correlation and the wireless link capacities. The proposed algorithm considers both factors, while minimizing the total energy consumed in the network. Moreover, the algorithm is asynchronous and amenable to fully distributed implementations, making it feasible for practical deployment in large-scale sensor networks. To the best of our knowledge, there does not exist any previous work that addresses the MEDG problem with data aggregation and wireless channel interference simultaneously, especially when a price-based strategy is employed to obtain a distributed algorithm to solve the problem. Finally, our recent research extends the proposed optimization framework to accommodate sensor networks with multiple sinks and arbitrary amount of data correlation [22].

References

[1] T. Clouqueur, V. Phipatanasuphorn, P. Ramanathan, and K. K. Saluja, "Sensor Deployment Strategy for Detection of Targets Traversing a Region," in *ACM Mobile Networks and Applications*, vol. 8, 2003, pp. 453–461.

[2] A. Mainwaring, J. Polastre, R. Szewczyk, and D. Culler, "Wireless Sensor Networks for Habitat Monitoring," in *Proc. of First ACM International Workshop on Wireless Sensor Network and Applications*, September 2002.

[3] B. Krishnmachari, D. Estrin, and S. Wicker, "The Impact of Data Aggregation in Wireless Sensor Networks," in *Proc. of the 22nd International Conference on Distributed Computing Systems*, 2002, pp. 575–578.

[4] P. Gupta and P. R. Kumar, "The Capacity of Wireless Networks," *IEEE Trans. on Information Theory*, vol. 46, no. 2, pp. 388–404, March 2000.

[5] B. Krishnamachri, D. Estrin, and S. Wicker, "Modelling Data-Centric Routing in Wireless Sensor Networks," in *Proc. of IEEE INFOCOM*, 2002.

[6] K. Kalpakis, K. Dasgupta, and P. Namjoshi, "Efficient Algorithms for Maximum Lifetime Data Gathering and Aggregation in Wireless Sensor Networks," *Computer Networks Journal*, 2002.

[7] A. Goel and D. Estrin, "Simultaneous Optimization for Concave Costs: Single Sink Aggregation or Single Source Buy-at-Bulk," in *Proc. of the 14th Symposium on Discrete Algorithms (SODA)*, 2003.

[8] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate Control in Communication Networks: Shadow prices, Proportional Fairness and Stability," in *Journal of the Operational Research Society*, vol. 49, 1998, pp. 237–252.

[9] S. H. Low and D. E. Lapsley, "Optimization Flow Control: Basic Algorithm and Covergence," in *IEEE/ACM Trans. on Networking*, vol. 7, 1999, pp. 861–874.

[10] Y. Xue, B. Li, and K. Nahrstedt, "Optimal Resource Allocation in Wireless Ad Hoc Networks: A Price-Based Approach," in *to appear in IEEE Transactions on Mobile Computing*, 2005.

[11] R. G. Gallager, P. A. Humblet, and P. M. Spira, "A Distributed Algorithm for Minimum-Weight Spanning Trees," in *ACM Transactions on Programming Languages and Systems*, vol. 5, no. 1, January 1983, pp. 66–77.

[12] H. D. Sherali and G. Choi, "Recovery of primal solutions when using subgradient optimization methods to solve Lagrangian duals of linear programs," in *Operations Research Letter*, vol. 19, 1996, pp. 105–113.

[13] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Prentice Hall, 1989.

[14] J. Czyzyk, M. Mesnier, and J. More, "The NEOS Server," in *IEEE Journal on Computational Science and Engineering*, vol. 5, 1998, pp. 68–75.

[15] J. H. Chang and L. Tassiulas, "Maximum Lifetime Routing in Wireless Sensor Networks," in *IEEE/ACM Transactions on Networking*, vol. 12, August 2004, pp. 22–31.

[16] B. Hamdaoui and P. Ramanathan, "Energy-efficient and MAC-aware Routing for Data Aggregation in Sensor Networks," in *Proc. of Sensor Network Operations*, 2004.

[17] F. Ordonez and B. Krishnamachari, "Optimal Information Extraction in Energy-Limited Wireless Sensor Networks," in *IEEE Journal on Selected Areas in Communications, special issue on Fundamental Performance Limits of Wireless Sensor Networks*, 2004.

[18] M. Johansson, L. Xiao, and S. Boyd, "Simultaneous Routing and Resource Allocation in CDMA Wireless Data Networks," in *Proc. of IEEE International Conference on Communications*, vol. 1, May 2003, pp. 51–55.

[19] P. V. Rickenbach and R. Wattenhofer, "Gathering Correlated Data in Sensor Networks," in *Proc. of DIALM-POMC '04: Proceedings of the 2004 Joint Workshop on Foundations of Mobile Computing*, 2004, pp. 60–66.

[20] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient Communication Protocol for Wireless Sensor Networks," in *Proc. of the Hawaii International Conference System Sciences*, January 2000.

[21] S. Lindsey and C. S. Raghavendra, "PEGASIS: Power Efficient GAthering in Sensor Information Systems," in *Proc. of the IEEE Aerospace Conference*, March 2002.

[22] K. Yuen, B. Li, and B. Liang, "Distributed Data Gathering in Multi-Sink Sensor Networks with Correlated Sources," to appear in *Proc. of IFIP Networking*, May 2006.