

Minimum Cost Hierarchical Architecture for Correlated Data Aggregation in Sensor Networks with Mobile Access

Tran Minh Tam, Hung Q. Ngo, Sungyoung Lee
Department of Computer Engineering
Kyung Hee University, South Korea, 446-701
{tmtam, nqhung, sylee}@oslab.khu.ac.kr

Miso Kim
Telecommunication R&D Center
Samsung Electronics, South Korea
meeso.kim@samsung.com

Abstract—We consider an energy-efficient data aggregation scheme for sensor networks with mobile access (SENMA), a design that opens the possibility of relieving low-power nodes from energy-consuming networking tasks. Specifically, we provide a novel distributed one-hop clustering algorithm to organize the network into a minimum cost hierarchical architecture (MCHA) for collecting all sensing data from the monitored area. For better load balancing and prolonged network lifetime, we exploit the joint effect of clustering and data correlation and provide a trade-off between communication cost and residual energy of sensor nodes. Our algorithm, which employs a message-passing algorithm for solving the min-sum optimization problem, is simple to implement and gives a near-optimal solution within a small number of iterations. Energy efficiency, scalability and robustness of the algorithm are also verified through extensive simulations.

I. INTRODUCTION

In this paper ¹, we consider a cluster-based approach for correlated data aggregation in sensor network with mobile access (SENMA) [1]. SENMA is a large-scale network of low-cost, wireless sensors that can report their sensing data directly to some mobile access points (or sinks, used interchangeably), which can be some kind of (manned/unmanned) aerial vehicles flying above a geographic area periodically, or a LEO satellite (see Fig. 1). This network paradigm is envisioned to inspire a wide range of applications in both civilian and military domains [2]. The propagation channels are likely line-of-sight, thus direct communications between sensors and sinks are possible even through a long distance [3]. There are some advantages which imply that SENMA is more favorable than the multihop ad-hoc network for such applications. The foremost one is energy efficiency [1], [4]: SENMA can relieve sensor nodes from the burden of operating MAC and routing functions. A set of selected sensors should be only waken up by the mobile sink and each active sensor

¹This research was supported by the MKE (Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the IITA (Institute of Information Technology Advancement) (IITA-2008-C1090-0801-0002). Also, this work was supported by the Korea Research Foundation Grant Funded by the Korean Government (MOEHRD) (KRF-2007-313-D00663). Corresponding author: Sungyoung Lee

should immediately return to its sleep mode after a successful transmission until the next visit of the mobile sink. The mobile sink, in contrast to low-cost sensors, is assumed to have its own power generator and sophisticated transceivers which provide multipacket reception (MPR) capability [5]. Obviously, direct sink access can significantly reduce the complexity of creating and maintaining a backbone for data forwarding (towards the sink). While synchronization at the physical layer among sensors under the multihop ad-hoc architecture can also be difficult, the presence of a strong beacon from the mobile sink in SENMA just simplifies the timing recovery and synchronization. For some applications the sensors require to know their locations, this can be conveniently facilitated by the mobile sink without running a localization algorithm which increases the computation overhead.

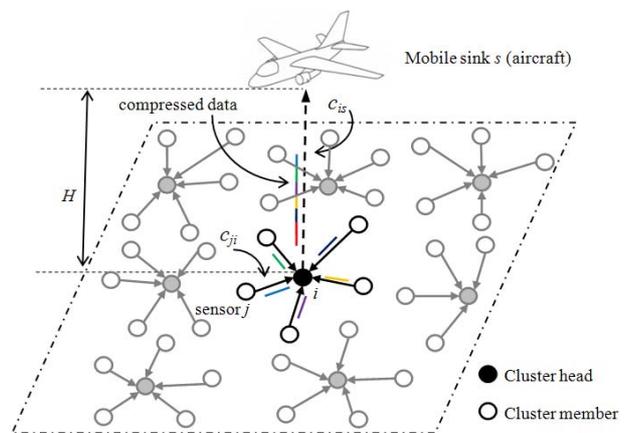


Fig. 1. Hierarchical architecture for gathering correlated data in a typical Sensor Network with Mobile Access (SENMA)

Most of the work so far focus on how to build an optimal tree structure for data aggregation towards a single fixed sink in a multihop ad-hoc sensor network. Unfortunately, these structures are no longer suitable for SENMA. On the other hand, clustering – a mechanism to organize the sensor network into a connected hierarchy of cluster heads (CHs) and member nodes – is a natural approach that supports sink

mobility and multiple sink network. Clustering can also aid in increasing the scalability, reducing the energy consumption, and extending the network lifetime [6] which implies the need for load balancing, sleep and transmission scheduling, efficient resource utilization, and in-network data fusion.

It is long known that a degree of spatial correlation always exists in the observed data of neighboring sensors which leads to significant waste of transmission energy due to inherent data redundancy. Data aggregation is used to eliminate such data redundancy by compressing the sensor measurements at intermediate nodes. The problem of data gathering and aggregation from a number of distributed sources with a certain correlation structure is generally considered as a joint-optimization of rate allocation and transmission structure [7], [8], [9], [10], [11], [12]. The rate allocation optimization problem is to find the minimum data encoding rate at each node, while ensuring the data collected by the sinks can be decoded to reconstruct the original data [13]. The transmission structure optimization problem is to find optimal routes from the source nodes to the sinks such that the total energy consumption incurred by transmitting all aggregated data is minimized. We can use either distributed source coding (e.g. Slepian-Wolf coding) [14] or joint-entropy coding with explicit communication [15] for data compression. Slepian-Wolf coding requires a *synchronous* communication model and a *complete knowledge* of the network structure at each source, for example, the distance between every pair of nodes. Therefore, an optimal hierarchical architecture using this coding scheme becomes very difficult to establish in practice due to its computation complexity. On the contrary, sensor node that uses joint-entropy coding with explicit communication can encode a data source by only the side information received from at least one of neighboring nodes without waiting for other belated information. Although joint entropy coding with explicit communication is suboptimal in terms of data compression, it is more practical for implementation. Rickenbach et al. [7] classified it into two distinct techniques: *self coding* and *foreign coding*. With self coding, a node is able to encode its own raw data only in the presence of side information from at least another node (Fig. 2(a)). In contrast, foreign coding is the technique in which raw data originating at one node is encoded by another node (Fig. 2(b)). We can easily observe that foreign coding achieves better compression rate in a *many-to-one architecture* (cluster-based data aggregation scheme) than self coding.

Our objective is to minimize the total communication cost such that a maximum number of successful data gathering cycles can be achieved until the first/last node dies. The novelty of our proposed algorithm is twofold. We first reformulate the problem of finding a Minimum Cost Hierarchical Architecture (MCHA) as a min-sum labeling problem. Specifically, we force the algorithm to search for the optimal labeling in a constrained search space containing valid configurations only. This allows us to reduce the time and message complexity in finding the optimal solution. Second, we use the recent modeling and computational methodology of factor graphs

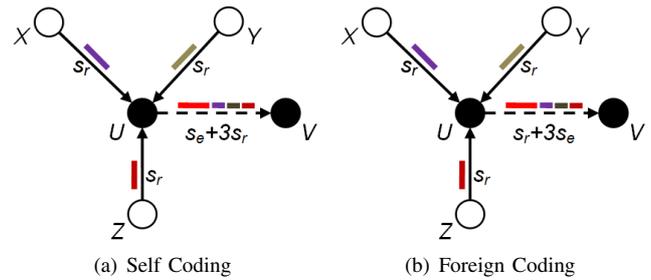


Fig. 2. Two techniques of joint entropy coding with explicit communication

and message-passings [16], which are very convenient for distributed realization in WSNs, to model and solve the min-sum labeling problem [17]. Each node in the network uses local information obtained from its neighborhood to calculate, update, and exchange messages with neighboring nodes only. Thus, no message routing mechanisms are invoked. The near-optimal set of cluster head (CH) nodes emerges after a few iterations using such simple and localized message-passing rules. Since the CH selection is in favor of nodes with higher residual energy to reduce the reclustering rate which is energy wasted, we provide an analysis on the trade-off between the residual energy and the communication cost. Simulation results show that the our algorithm can quickly achieve a good approximation of the minimum cost found by a centralized Mixed Integer Linear Programming (MILP) algorithm using CPLEX optimization package. Through extensive simulations, we can affirm that MCHA is very scalable, robust, and adaptive, as a result of the highly localized nature of the algorithm. To the best of our knowledge, this is the first time the cluster-based architecture for correlated data aggregation in SENMA is formulated as a min-sum constraint-satisfaction problem and efficiently solved using the message-passing approach.

The rest of this paper is organized as follows. After discussing about the related work and their limitations in Section II, we first formulate the correlated data aggregation problem as an NP-hard min-sum optimization problem in Section III. Next, we describe the simplified message-passing protocol for the min-sum labeling in Section IV. We evaluate the proposed algorithm through simulations, and compare its effectiveness to other approaches in Section V.

II. RELATED WORK

Hierarchical network architecture for data aggregation has been the focus of much research work. Using clustering approach, the authors of LEACH [6] and HEED [18] mainly focus on reducing the overhead and improving the scalability but do not consider the effect of spatial data correlation. Wang et al. [12] propose a heuristic algorithm for constructing clusters of nodes so that Slepian-Wolf coding can be performed locally. The advantage of this approach is the reduction of computation complexity and relay failures, which affect on the data reconstruction at the sink. This protocol only requires a localized knowledge of the network structure and is totally distributed. However, it is based on a sequential greedy

method regarding the clustering problem as a well-known Minimum Weight Set Covering problem, which is NP-hard. A combinatorial explosion of computations is *unavoidable* to find the best value of a heuristic function based upon the conditional entropies of a set of nodes. Moreover, the residual energy factor is not taken into account, thus some low-power nodes can be chosen as cluster heads. Cluster-based data aggregation for SENMA has been recently considered in [19] and [2]. In [19], the authors consider the deployment of two types of sensor nodes such that a minimum cost heterogeneous sensor network is built with a lifetime constraint. However, data correlation is not considered in this work. The latest work [2] done by Lotfinezhad propose a Cluster-based Data Collection scheme for sensor network with Direct Sink Access. In this scheme, a low-overhead and simple MAC conceptually similar to ALOHA is provided and evaluated under an Energy-Efficient Clustering (EEC) framework derived from [20]. The focus of [2] is thus not minimum cost hierarchical architecture but transmission scheduling at MAC layer.

III. PROBLEM FORMULATION

A. Network Model

Consider N stationary sensors (e.g. MICA mote) are randomly and uniformly deployed and left unattended in a square area. They are represented as vertices connected to each other in an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where a vertex in \mathcal{V} represents a sensor and an edge in \mathcal{E} corresponds to a link between two nodes. Every node can save energy by using a low power level for short-range (intra-cluster) communications but a higher level for transmitting the data through a long distance to reach to mobile sink. The weight of an edge (i, j) in the graph is proportional to the communication cost c_{ij} between node i and node j , which is a function of energy spent for transmitting and receiving one data unit from node i to node j . Nodes i and j are neighbors if they are connected by an edge, i.e. $(i, j) \in \mathcal{E}$. We define an *open neighbor set* of node i is defined as $\mathcal{N}(i) = \{j | (i, j) \in \mathcal{E}\}$, a *closed neighbor set* of node i as $\mathcal{N}[i] := \mathcal{N}(i) \cup \{i\}$, and $\mathcal{N}(i) \setminus j$ denotes the set obtained by excluding j from $\mathcal{N}(i)$. Since a cross-layer design is out of the scope of this paper, we assume a contention and error-free MAC protocol. However, it should be clear that MAC is much more difficult to handle under the flat multihop ad-hoc architecture than SENMA, and is likely to incur a more significant overhead in energy consumption.

B. Cost Model

The communication cost c_{ij} , which represents the energy consumption for transmitting one bit of data from a sensor i to its neighbor j , can be estimated as $c_{ij} = \varepsilon_A d_{ij}^\alpha + \varepsilon_E$, where d_{ij} is the Euclidean distance between i and j , α is the power attenuation factor ($2 \leq \alpha \leq 4$), while $\varepsilon_A, \varepsilon_E$ are constants ($\varepsilon_A, \varepsilon_E > 0$). Constant ε_E accounts for the energy needed to run electronic circuits at transmitter and receiver and minimal signal strength for correct signal reception. Constant ε_A accounts for the energy needed to run the power amplifier at the transmitter [6]. The mobile sink s is assumed to fly over

the network at a height H far enough such that the distance between a CH i and s can be approximated by H [2] ($d_{is} = H$). Since the radio propagation between cluster heads and the mobile sink (in our case, the aircraft) is apparently line-of-sight, a path loss exponent of 2 – the free space model [6] – is considered for both internode communication and direct sink access ($\alpha = 2$). The cost for uploading a data unit from a CH i to the mobile sink s is thus $c_{is} = \varepsilon_A H^2 + \varepsilon_E$.

C. Correlation Model

In reality, the data observed by a sensor is correlated with the data of its neighbor nodes according to a specific structure. Let the parameter η_{ij} be the compression rate, the reduced data rate by compressing raw data of node i at node j using information available at node j (i.e. foreign coding). This parameter depends on the distance d_{ij} between i and j . The sensing data of the monitored area is assumed to be Gaussian such that the correlation between every pair of nodes can be modeled by a covariance matrix [21]. We use the Power Exponential model for the correlation coefficient such that the reduced data rate can be assigned by the following expression:

$$\eta_{ij} = 1 - e^{-\gamma d_{ij}^2} \quad (1)$$

where γ is a constant indicating the degree of correlation. This correlation model can capture a wide variety of physical phenomenon in practice [22]. Each sensor i is supposed to estimate η_{ij} using distance estimation or through several message exchanges of sensing data [9].

D. Problem Setup

The total cost in a cluster-based data gathering problem is the sum of intra-cluster communication cost for sending raw data from cluster members to their corresponding CHs, and the cost for relaying the compressed data from CHs to sink. Our minimum cost hierarchical architecture problem can be regarded as a *min-sum labeling* problem: *Identifying a subset of nodes in the network to label as CHs, then assigning each of the remaining nodes to the CH with minimum transmission cost, so that the total transmission cost of the data aggregation application is minimized.*

The label to which a node is assigned can be considered as a latent variable. Let $X := \{x_1, x_2, \dots, x_N\}$ be a vector of N such hidden variables, in which, for each i , x_i takes on values (node IDs) in $\mathcal{N}[i]$. The estimated cost ζ_i for a sensor node i to deliver one data unit to the sink is supposed to be c_{is} if i acts as a CH; or c_{ij} for a transmission from i to its CH j within one hop of i , plus the *relay cost* $\eta_{ij}c_{js}$ for sending the encoded data from CH j to the mobile sink s :

$$\zeta_i(x_i) = \begin{cases} c_{is} & \text{if } x_i = i \\ c_{ij} + \eta_{ij}c_{js} & \text{if } x_i = j, j \in \mathcal{N}(i) \end{cases} \quad (2)$$

where η_{ij} is the reduced data rate obtained from Eq. (1).

In a cluster-based SENMA, the CHs tend to run out of energy more quickly because they must transmit a large amount of data through a long distance. In order to make

the network load-balanced, we need to re-elect other high-energy nodes to act as CHs periodically. However, re-election (or re-clustering, used interchangeably) is a waste of energy, thus we need to reduce the reclustering rate. By scaling the transmission cost function with the *relative* residual energy of the related nodes as below, we can provide a trade-off between node residual energy and communication cost:

$$\zeta_i(x_i) = \begin{cases} \frac{e_0}{e_i} c_{is} & \text{if } x_i = i \\ \frac{e_i}{e_j} (c_{ij} + \eta_{ij} c_{js}) & \text{if } x_i = j, j \in \mathcal{N}(i) \end{cases} \quad (3)$$

where e_0 is the initial energy of a node; e_i, e_j are the residual energy of node i and node j , respectively.

Let the N-tuple $\mathbf{x} := (x_1, x_2, \dots, x_N)$ denote the *configuration* (or *labeling*) of the whole network. Since the system is specified via its configuration, this approach is also known as *behavioral modeling*. The N-tuple \mathbf{x} can be a *valid* or *invalid* configuration [23]. For example, if node i selects j as its CH (i.e., $x_i = j$), but node j is not correctly labeled as a CH (e.g., $x_j = k \neq j$), then this is apparently an invalid configuration. To enforce valid configurations between the label x_i of sensor i and the labels of its 1-hop neighboring nodes, denoted as $x_{1:I}^i$, with $I = |\mathcal{N}(i)|$, a constraint function $\theta_i(x_i, x_1^i, \dots, x_I^i)$ is used. For the min-sum configuration problem, this function gives a penalty of 0 or ∞ for a valid or invalid configuration respectively, defined as follows:

$$\theta_i(x_i, x_1^i, \dots, x_I^i) := \begin{cases} \infty, & \text{if } x_i \neq i \text{ but } \exists i' \in \mathcal{N}(i) : x_{i'}^i = i \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

It is worth noting that the constraint functions just serve as a mathematical modeling of the problem at hand; they do not put any burdens on the computation of the algorithm. The problem of choosing a minimum-cost hierarchical architecture for data gathering with in-network aggregation now becomes the problem of finding the min-sum labeling among the *valid* configurations, defined as:

$$\mathbf{x}_{opt} := \arg \min_{\mathbf{x}} \left[\sum_{i \in \mathcal{V}} \zeta_i(x_i) + \sum_{i \in \mathcal{V}} \theta_i(x_i, x_1^i, \dots, x_I^i) \right] \quad (5)$$

It is known that exactly minimizing the overall cost is computationally intractable, since a special case of this problem is the NP-hard k-mean problem in data clustering; for large problem we can only find approximate solutions which are heuristic in nature. We propose a new approach for finding a near-optimal solution by recursively applying the min-sum message-passing algorithm [16].

IV. MESSAGE-PASSING DISTRIBUTED PROTOCOL FOR BUILDING MINIMUM COST HIERARCHICAL ARCHITECTURE

In this section, we describe the simplified message-passing rules of our Message-Passing Distributed protocol. Factor graphs [16] can be used to represent a complicated global function, which can be factored into simpler “local” functions, each of which depends on a subset of the variables. In a

factor graph, message-passing algorithms can compute, either exactly or approximately, various function marginalization and maximization using simple message passing rules. More details on factor graph derivation for the min-sum labeling problem given in Eq. 5 and message simplification can be found in our prior work [17]. The two types of messages exchanged between the sensor nodes in the network graph \mathcal{G} are:

- Request message $\mu_{x_i \rightarrow \theta_j}$ sent from sensor i to its neighbor $j \in \mathcal{N}[i]$, reflects the accumulated level of *suitability* for sensor i to select neighbor j as its CH, taking into account other neighboring CH candidates j' of i .

$$\mu_{\theta_j \rightarrow x_i} = \max \left[0, \mu_{x_j \rightarrow \theta_j} + \sum_{j' \in \mathcal{N}(j) \setminus i} \min \left(0, \mu_{x_{j'} \rightarrow \theta_j} \right) \right] \quad (6)$$

- Reply message $\mu_{\theta_i \rightarrow x_j}$ sent from sensor i to its neighbor $j \in \mathcal{N}[i]$, reflects the accumulated level of *willingness* of sensor i to act as a CH for sensor j , taking into account the requests from other neighbors j' of i .

$$\mu_{x_i \rightarrow \theta_j} = \zeta_i(j) - \min_{j' \in \mathcal{N}[i] \setminus j} \left[\zeta_i(j') + \mu_{\theta_{j'} \rightarrow x_i} \right], \forall j \in \mathcal{N}[i] \quad (7)$$

Each request/reply message contains a single number such that it allows a node to marshal all the request and reply messages into a vector message *COMPACT* and send to all of its neighbors by a single broadcast. The messages can be initialized arbitrarily. In our implementation we initialize them to zeros.

The proposed one-hop clustering algorithm is fully distributed and can be efficiently implemented in real sensors because it involves only simple computations using information available via message broadcast without any routing mechanism. The near optimal set of CHs emerges from this recursive message-passing procedure. At any iteration, each node can evaluate its intermediate CH candidate by identifying the sensor *ID* in its *closed* neighbor set given the expression:

$$CH_i = \arg \min_{j \in \mathcal{N}[i]} [\zeta_i(j) + \mu_{\theta_i \rightarrow x_j}] \quad (8)$$

Each node begins to run the algorithm with a Cluster Head Election process and this process is terminated when the maximum number of iterations *maxIter* is reached. This is a key parameter that needs to be carefully selected in real implementation, since the more number of recursions, the better approximation of the optimal clustering, at the cost of more messages to be exchanged. We found through simulations that *maxIter* = 5 is a reasonable upper bound. After the Cluster Head Election period, each CH broadcasts an advertisement to its neighboring nodes. A node, which does not become a CH, then chooses and joins the neighboring CH with least cost (see Fig. 3).

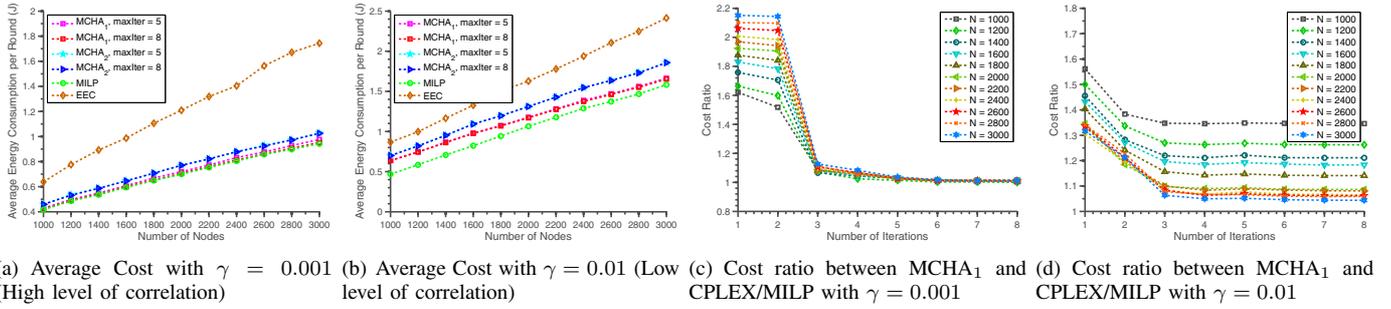


Fig. 4. Cost (Energy Consumption per Round) Analysis of solutions found by MCHA₁, MCHA₂, CPLEX/MILP and EEC

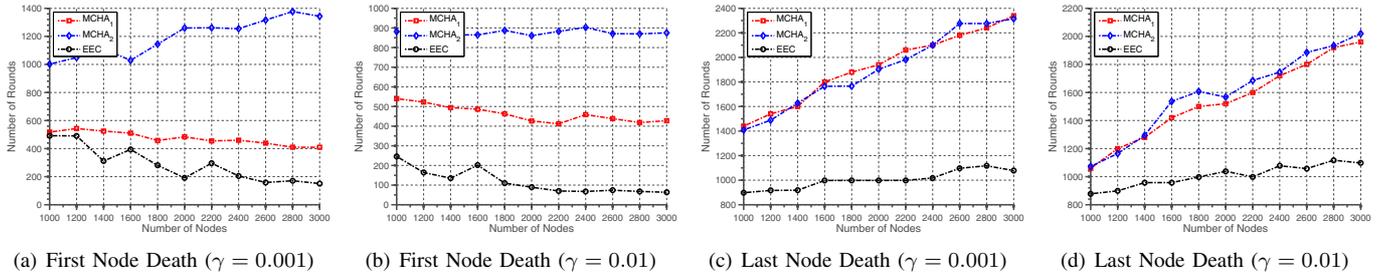


Fig. 5. Network Lifetime ($maxIter = 5$ for MCHA)

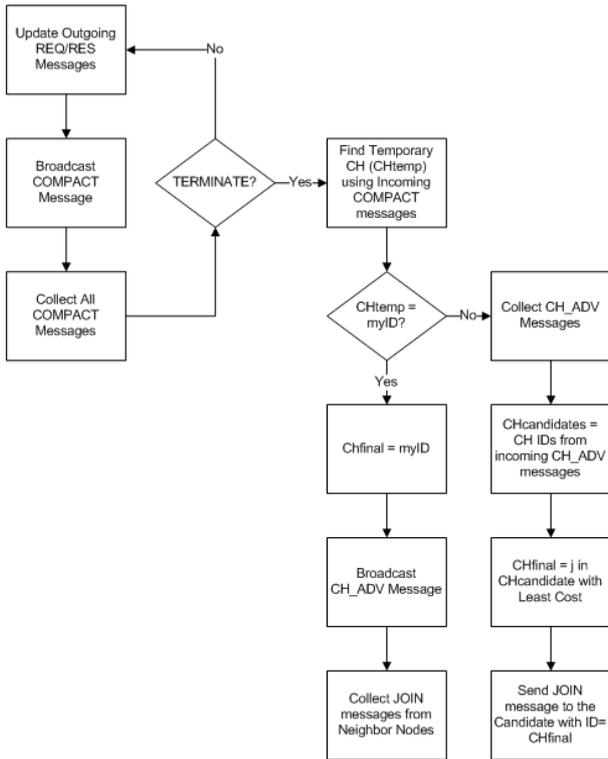


Fig. 3. Protocol Flowchart

V. PERFORMANCE EVALUATION AND ANALYSIS

In this section, we provide our simulation results with network setups involving a large number of sensor nodes distributed uniformly in a wide square area ($1000 \times 1000m^2$)

TABLE I
SIMULATION PARAMETERS

Type	Parameter	Value
Network	Network Size	$1000 m \times 1000 m$
	Flight Altitude H	$250 m$
	Initial Energy e_0	$2 J$
	Radio Range r	$25 m$
	Broadcast Packet Size	$25 bytes$
	Raw Data Packet Size	$125 bytes$
Radio Model	ϵ_A	$10 pJ/bit/m^2$
	ϵ_E	$50 nJ/bit$
	$maxIter$	5
	$reclusteringRate$	$50 rounds$

based on a range of parameters provided in Table I. Two clustered network configurations are considered depending on which cost function is used:

- 1) MCHA₁ for the cost function in Eq. (2) – transmission cost only.
- 2) MCHA₂ for the cost function in Eq. (3) – transmission cost scaled with node residual energy.

We first analyze the average total cost (i.e. energy consumption) per round. A round is defined as the time period during which sensing data from all sensor nodes is collected by the mobile sink once. We compare the cost found by our distributed MCHA algorithm with the optimal cost found by the centralized Mixed Integer Linear Programming (MILP) method using the TOMLAB/CPLEX optimization package and the cost of the Energy Efficient Clustering (EEC) protocol proposed in [20]. Briefly described, EEC finds clusters of nodes in which a node checks if it is a volunteer CH or not probabilistically; if a node is not a CH, it becomes a member

node of the closest CH within a maximum of k hops. As we see in Fig. 4, the performance of MCHA tracks well with that of the centralized, optimal MILP method and far better than that of EEC. The cost of MCHA₁ approximates the optimal centralized solution better with high degree of data correlation (Fig. 4(a)) but the approximation cost ratio is also very good even with a low degree of correlation (Fig. 4(b)).

In order to determine the threshold $maxIter$, we analyze the cost ratio between our message-passing algorithm MCHA₁ and the centralized optimal solution by TOMLAB/CPLEX package. Through 4(c) and 4(d), it can be seen that $maxIter = 5$ is a reasonable upper bound for the minimum estimated cost regardless of other parameters such as N or γ . Therefore, we use this value in the subsequent analysis on network lifetime of the data-aggregation applications in cluster-based SENMA.

Through Fig. 5(a) and 5(b), we see that the network lifetime (i.e. the number of rounds) of MCHA₂ is increased significantly compared with that of MCHA₁ in the First-Node-Death case although the energy consumption per round also increases slightly as shown in Fig. 4, due to the effect of energy scaling which leads to a suboptimal solution. This can be explained that a trade-off between the communication cost and the node residual energy, in which a set of relatively high-energy CHs is elected through the message-passing recursions, is taken into account for MCHA₂. Furthermore, it is easily observed that the performance of our approaches is much better than that of EEC, which does not consider the effect of data correlation. Fig. 5(c) and 5(d) also show us the network lifetimes of MCHA₁ and MCHA₂ are almost the same and much longer than that of EEC in the Last-Node-Death scenario.

By carrying out these simulations with the number of nodes varied from 1000 to 3000 and two extreme cases of data correlation ($\gamma = 0.001 \mid 0.01$), we show that our distributed MCHA is very scalable and best suitable for monitoring a vast area in large-scale SENMA.

VI. CONCLUSION AND FUTURE WORK

In this paper, we introduced a Minimum Cost Hierarchical Architecture for Correlated Data Gathering (MCHA) in wireless sensor networks with mobile access. Numerical simulations show that network lifetime can be increased significantly by our cluster-based data aggregation approach, which is scalable, fully distributed and easy to implement in the real environment. We are developing a distributed reclustering mechanism for better load balancing so as to further enhance the performance of the proposed algorithms.

REFERENCES

- [1] L. Tong, Q. Zhao, and S. Adireddy, "Sensor networks with mobile agents," in *Proc. IEEE MILCOM'03*, Oct. 2003, pp. 310–321.
- [2] M. Lotfinezhad, B. Liang, and E. S. Sousa, "Adaptive Cluster-Based Data Collection in Sensor Networks with Direct Sink Access," vol. 7, no. 7, pp. 884–897, Jul. 2008.
- [3] P. Venkatasubramanian, S. Adireddy, and L. Tong, "Sensor networks with mobile access: Optimal random access and coding," vol. 22, no. 6, pp. 1058–1068, Aug. 2004.
- [4] G. Mergen, Q. Zhao, and L. Tong, "Sensor Network With Mobile Access: Energy and Capacity Considerations," vol. 4, no. 1, pp. 4–15, Nov. 2006.
- [5] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "Stability properties of slotted ALOHA with multipacket reception capability," vol. 14(1), pp. 41–54, Feb. 2006.
- [6] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," vol. 1(4), pp. 660–670, Oct. 2002.
- [7] P. von Rickenbach and R. Wattenhofer, "Gathering correlated data in sensor networks," in *Proc. ACM DIALM-POCM'04*, Philadelphia, Pennsylvania, USA, Oct. 2004.
- [8] M. Lotfinezhad and B. Liang, "Effect of partially correlated data on clustering in wireless sensor networks," in *Proc. IEEE SECON'04*, Oct. 2004, pp. 172–181.
- [9] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "Networked Slepian-Wolf: theory, algorithms, and scaling laws," vol. 51(12), pp. 4057–4073, Dec. 2005.
- [10] —, "Network correlated data gathering with explicit communication: Np-completeness and algorithms," vol. 14(1), pp. 41–54, Feb. 2006.
- [11] J. Liu, M. Adler, D. Towsley, and C. Zhang, "On optimal communication cost for gathering correlated data through wireless sensor networks," in *Proc. ACM MobiCom'06*, 2006, pp. 310–321.
- [12] P. Wang, C. Li, and J. Zheng, "Distributed data aggregation using clustered slepian-wolf coding in wireless sensor networks," in *Proc. IEEE ICC'07*, Glasgow, Scotland, Jun. 2007, pp. 3616–3622.
- [13] K. Yuen, B. Liang, and L. Baochun, "A distributed framework for correlated data gathering in sensor networks," vol. 57, no. 1, pp. 578–593, Jul. 2008.
- [14] D. Slepian and J. Wolf, "Noiseless Coding of Correlated Information Sources," vol. 19, no. 4, pp. 471–480, Jul. 1973.
- [15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: John Wiley and Sons, Inc., 2006.
- [16] F. R. Kschischang, B. Frey, and H. Loeliger, "Factor graphs and the sum-product algorithm," vol. 47, no. 2, pp. 498–519, Nov. 2001.
- [17] H. Q. Ngo, T. M. Tam, Y. K. Lee, and S. Y. Lee, "A message-passing approach to min-cost distributed clustering in wireless sensor networks," in *Proc. IEEE ATC '08*, Hanoi, Vietnam, Oct. 2008, pp. 174–178.
- [18] O. Younis and S. Fahmy, "HEED: A Hybrid, Energy-Efficient, Distributed clustering approach for ad hoc sensor networks," vol. 3, no. 4, Oct.-Dec. 2004.
- [19] V. P. Mhatre, C. Rosenberg, D. Kofman, R. Mazumdar, and N. Shroff, "A Minimum Cost Heterogeneous Sensor Network With a Lifetime Constraint," vol. 4, no. 1, pp. 4–15, 2005.
- [20] S. Bandyopadhyay and E. Coyle, "An energy efficient hierarchical clustering algorithm for wireless sensor networks," in *Proc. IEEE INFOCOM'03*, Apr. 2005, pp. 1713–1723.
- [21] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "On network correlated data gathering," in *Proc. IEEE INFOCOMM'04*, vol. 4, Hongkong, Mar. 2004, pp. 2571–2582.
- [22] M. C. Vuran and I. F. Akyildiz, "Spatial correlation-based collaborative medium access control in wireless sensor networks," vol. 14, no. 2, pp. 316 – 329, Apr. 2006.
- [23] J. C. Willems, "Models for dynamics," in *Dynamics Reported, Volume 2*, U. Kirchgraber and H. O. Walther, Eds. New York: Wiley, 1989, pp. 171–269.