Energy-Efficient Clustering in Wireless Sensor Networks with Spatially Correlated Data

Ali Dabirmoghaddam   Majid Ghaderi   Carey Williamson

Department of Computer Science, University of Calgary

{adabirmo, mghaderi, carey}@cpsc.ucalgary.ca

Abstract—In this paper, we show how spatial correlation in data can be exploited to reduce energy consumption in a wireless sensor network. We analytically characterize the optimal cluster size, and then use a greedy clustering algorithm to study approximate solutions to the optimal data gathering problem.

I. INTRODUCTION

A *Wireless Sensor Network* (WSN) consists of a large set of sensor nodes that cooperate to monitor environmental conditions (e.g., temperature, precipitation, radioactivity) in a given geographic area. WSNs are often designed for long-term operation in remote unattended environments, despite the limited battery capacity of the wireless sensor nodes.

Since data transmission is an energy-intensive task, energy-aware data gathering techniques are used to extend the lifetime of the WSN. An effective way to conserve energy is to avoid reporting redundant data that occurs due to the spatial correlation between nearby readings. However, *minimum cost network correlated data gathering* is NP-complete [1].

Clustering is a well-established technique for reducing data collection costs in WSNs [2]. In these schemes, sensor nodes are grouped into disjoint sets each managed by a designated *Cluster-Head* (CH). Cluster-based architectures help overcome scalability issues (e.g., routing and communication costs) in large WSNs, and also reduce the volume of data transmitted to the sink, by suppressing local data redundancy.

In this paper, we carefully analyze the impact of data-dependent clustering on network-wide energy consumption. In particular, we provide a simple model to study the joint effect of cluster size and distance from the sink on reducing total network energy consumption for spatially correlated data. We also use a *"Greedy Clustering Algorithm"* to approximate the optimal data gathering tree in such WSNs.

II. PROBLEM FORMULATION

Consider a planar region in which sensor nodes are randomly and uniformly distributed with average density ρ. Sensors observe some spatial stochastic process, quantize their observations, and transmit them to the sink, either directly (single-hop) or via other sensor nodes (multi-hop). For successful data transmission, a minimum received power level γ is required. We assume a large-scale fading channel between each transmitter and receiver, in which the received power is inversely proportional to the square of the distance between the transmitter and the receiver.

Suppose that an arbitrary sensor node at distance L from the sink is being considered as a CH (see Fig. 1). How large should its (circular) cluster be? For a disk-shaped cluster of radius R, the expected number of sensors in the cluster is $N = \rho \pi R^2$.

In our analysis, we consider two communication schemes (direct and cluster-based) and compute the total transmission power required to report all observations to the sink.
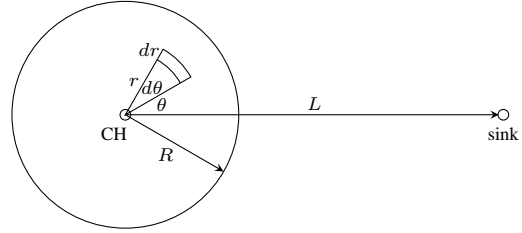

Fig. 1.   A circular cluster of radius R at distance L from the sink.

In direct communication, each sensor in the disk individually encodes its observation and transmits the message directly to the sink. Let b_1 denote the number of bits required to encode each observation. For a sensor node at polar coordinate (r, θ), as shown in Fig. 1, the transmission cost for sending a message to the sink is $P(r, \theta) = \gamma b_1 d^2(r, \theta)$, where $d(r, \theta)$ is the Euclidean distance from (r, θ) to the sink. The total transmission power P_d consumed by all sensors is obtained by integrating over the disk as follows:

$$P_d = \int_0^R \int_0^{2\pi} \gamma b_1 \left(r^2 + L^2 - 2rL\cos\theta \right) \rho r d\theta dr$$
$$= 2\pi\gamma\rho b_1 \left(\frac{L^2 R^2}{2} + \frac{R^4}{4} \right), \tag{1}$$

where the factor $\rho r d\theta dr$ represents the expected number of sensor nodes in the differential shown in Fig. 1.

In cluster-based communication, each sensor sends its quantized observation to the CH. The CH (losslessly) compresses all N messages and transmits the compressed version to the sink. Denoting the compressed message size by b_N, the total transmission cost is:

$$P_c = \int_0^R \int_0^{2\pi} \left(\gamma b_1 r^2 \right) \rho r d\theta dr + \gamma b_N L^2$$
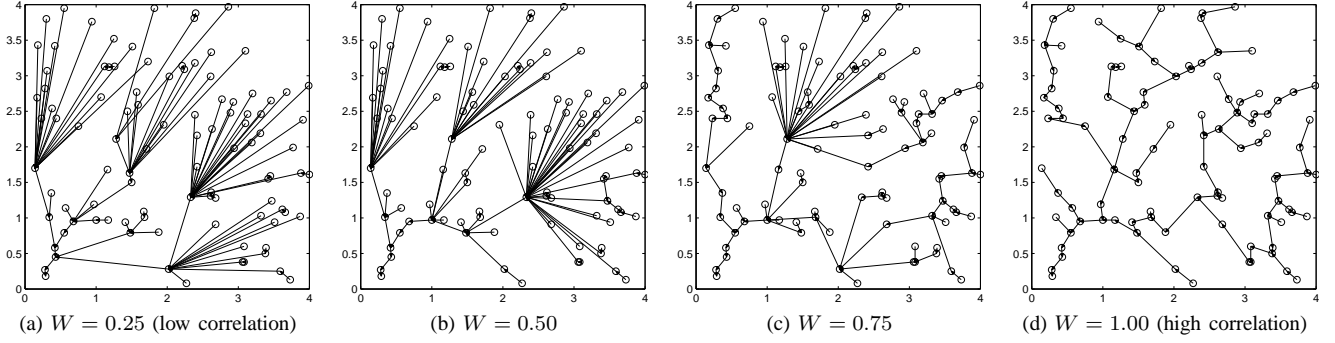$$= 2\pi\gamma\rho b_1 \left(\frac{R^4}{4} \right) + \gamma b_N L^2. \tag{2}$$

Fig. 2. Greedy clustering algorithm results for different degrees of data correlation.

(a) $W = 0.25$ (low correlation) (b) $W = 0.50$ (c) $W = 0.75$ (d) $W = 1.00$ (high correlation)

We focus on b_N, and its dependence on data correlation. We assume that the individual sensor observations follow a zero mean normal distribution with variance σ^2. The cluster, hence, can be described as a Gaussian random field with mean zero and *correlation matrix* $\Sigma = [\sigma_{ij}]_{N \times N}$, where $\sigma_{ij} = \sigma^2 W^{d_{ij}^2}$ denotes the data correlation between sensors i and j separated by distance d_{ij}. The parameter $0 \le W \le 1$ adjusts the correlation degree. In a multi-variate Gaussian distribution, the joint entropy of the set of quantized sources $S = \{S_1, \cdots, S_N\}$ is given by [3]:

$$H(S^D) \approx \frac{1}{2} \log_2 \left(\frac{\pi e}{6D} \right)^N |\Sigma| , \qquad (3)$$

where D is the (tolerable) target distortion, and $|\Sigma|$ denotes the determinant of the covariance matrix. Equation 3 can be used to compute the minimum number of bits required to represent an individual sensor reading (b_1) and the entire cluster after compression (b_N) with respect to a pre-specified target distortion D, respectively.

III. OPTIMAL CLUSTER SIZE

We evaluate the impact of the cluster size on the total energy consumption using the *"energy ratio"* metric $E = P_c / P_d$. To compute our numerical examples of energy ratio, we place a CH at distance L from the sink and dynamically expand the cluster radius R for a given node density ρ.
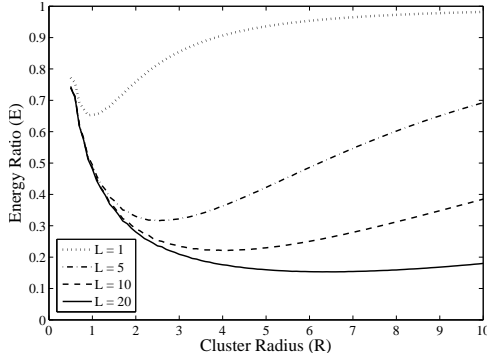

Fig. 3. The joint effect of the cluster size and the distance from the sink on the energy ratio (E) for $\rho = 6.25$ and $W = 0.75$.

Fig. 3 shows that, for a given distance L, there is an optimal cluster radius for which the energy ratio is minimized. The

farther the cluster is from the sink, the larger the optimal radius is. In other words, the optimal cluster size is not uniform throughout the network. This insight is important, since many energy-aware clustering schemes, such as LEACH and EEHC, assume homogeneous-sized clusters that are uniformly distributed in the WSN [2].

IV. THE GREEDY CLUSTERING ALGORITHM

To verify our results from the previous section, we simulated a centralized greedy clustering algorithm to find near-optimal solutions for the clustering problem. A distributed version of this algorithm remains as future work.

The greedy algorithm selects the WSN node with the highest cost to reach the sink, and then assigns it to the candidate CH for which the overall energy reduction is largest. The node joins the best candidate cluster and is marked. The algorithm iteratively performs this task until all the nodes in the WSN are marked.

In our simulation, we uniformly scatter 100 sensor nodes in a WSN with sink at (0,0). The output of our greedy algorithm on this network is depicted in Fig. 2. Note that the physical placement of WSN nodes is the same in all experiments; only the data correlation degree changes.

Non-uniform cluster sizes are clearly evident in Fig. 2a. Clusters farther from the sink are larger in radius, consistent with the results in Fig. 3. For higher degrees of correlation, the number of visually distinct clusters reduces, and there is greater affinity to proximal nodes. In the extreme case of Fig. 2d with $W = 1.0$, the resulting topology resembles a Minimum Spanning Tree (MST) for the network, though it still differs from (and is more energy-efficient than) the MST.

The diverse behaviours in Fig. 2 show the important effect of spatial data correlation on optimal cluster formation.

REFERENCES

[1] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "On Network Correlated Data Gathering," in *IEEE INFOCOM*, March 2004, pp. 2571–2582.
[2] A. Abbasi and M. Younis, "A Survey on Clustering Algorithms for Wireless Sensor Networks," *Computer Communications*, vol. 30, no. 14-15, pp. 2826–2841, 2007.
[3] A. Scaglione and S. Servetto, "On the Interdependence of Routing and Data Compression in Multi-hop Sensor Networks," *Wireless Networks*, vol. 11, no. 1-2, pp. 149–160, 2005.