# DATA CENTRIC MULTI-SHIFT SENSOR SCHEDULING FOR WIRELESS SENSOR NETWORKS

*Jialin Zhang, Yu Hen Hu*

University of Wisconsin Madison
Electrical and Computer Engineering
1415 Engineering Drive, Madison, WI 53706

## ABSTRACT

A multi-shift sensor scheduling method is proposed to extend the operating lifespan of a wireless sensor network. Sensor nodes in the WSN are partitioned into N subnetworks and the operating schedule is partitioned into N shifts of equal duration. Exploiting spatial correlations among sensor nodes, data collected using each subnetwork can well approximate the data collected using original sensor network. Each subnetwork also form a connected component to ensure proper data collection. This task is formulated as a NP-hard constrained subset selection problem. A polynomial time heuristic algorithm leveraging breath-first search and subspace approximation is proposed. Simulations using a real world data set demonstrate superior performance and extended lifespan of this proposed method.

***Index Terms—*** Wireless sensor networks, node scheduling, data coverage, connectivity

## 1. INTRODUCTION

Wireless sensor network (WSN) is an emerging cyber physical system that gains much attention over the past decade [1]. The practical usefulness of a WSN not only depends on its performance but also on its operating lifespan. Prolonged operating lifespan lowers the maintenance cost and reduces downtime.

In a densely deployed WSN where numerous redundant sensor nodes are deployed, not all sensor nodes need to be active at all times. Instead, taking advantage of the spatial correlation among data sampled at neighboring sensor nodes, many sensor nodes may be switched off to low power sleep mode without affecting overall quality of sensing.

Previously, several *sensor node sleep scheduling* algorithms have been reported [2, 3, 4, 5, 6, 7, 8]. Among them, a small number of algorithms [9, 8, 10] eluded to a round-robin style multi-shift schedule to allow a group of sensor nodes taking turns to engage sensing tasks. However, the criterion these algorithms used to partition sensor nodes into different shifts are mostly based on physical locations of sensors which is not adaptive to dynamic, time varying sensing environment.

In this work, sensor nodes in the WSN are partitioned into N subnetworks and the operating schedule is partitioned into N shifts of equal duration. Each subnetwork only operates on one duration. A data centric performance criterion is adopted to optimally partition sensor nodes into different shifts. In particular, it is stipulated that data collected by the subset of sensor nodes in each shift should be sufficient for predicting the data would-be collected by those sleeping sensor nodes. This data recovery criterion can be easily adapted to heterogeneous sensing ranges and time varying sensing channel conditions and hence is inherently superior to the geographic coverage criterion. In addition, among the subset of sensor nodes assigned to the same shift, it is crucial to ensure these sensor nodes form a connected network so that sensed data could be forwarded to a fusion center.

In this work, the WSN lifetime extension problem is formulated as a constrained subset selection problem: partition the set of sensor nodes into $N$ subsets such that (i) the data collected by active nodes in each subset can best approximate the missing data from other sleeping nodes; and (ii) the connectivity of each subnetwork is maintained. We showed that the complexity of this problem is NP-hard. A polynomial time heuristic algorithm leveraging breath-first search and subspace approximation is proposed. Simulation using real world data set demonstrate superior performance and extended lifespan of this proposed method.

The rest of this paper is organized in the following way. In section 2, the background of the problem is introduced. In section 3, the sensor node scheduling problem is formulated. In section 4, the connectivity constrained sensor network partitioning algorithm is developed. In section 5, simulation set up and results are presented.

## 2. WSN NODE SCHEDULING PROBLEM

The WSN node scheduling problem consists of two components: performance and connectivity.

### 2.1. Data Coverage

When a subset of sensor nodes fall back to sleep mode, the overall sensing performance may be compromised compared to the situation when all sensor nodes are active. In this work, we use data coverage model to quantify the amount of performance degradation due to sensor sleeps. It is widely accepted that in a densely deployed WSN, data captured by neighboring sensors are highly correlated. Exploiting this spatial correlation, the sensor readings of sleep nodes may be recovered from those of the active nodes [11, 8]. Obviously, such a recovery often is not error free. Therefore, it is appropriate to use the data coverage error as a performance criterion of a partitioned WSN.

Denote $\mathbf{Y}$ to be a $t \times M$ matrix representing the sensor data collected from $M$ active sensor nodes within a particular shift of schedule over time indices $\{1, 2, \cdots, t\}$. Denote $\mathbf{X}$ to be a $t \times k$ matrix representing the sensor data that would be collected by the $k$ sleep sensor nodes during the same time duration. Our goal is to predict the $\mathbf{X}$ matrix based on the given $\mathbf{Y}$ matrix. This is possible if there are high spatial correlation among the reading of sensor nodes in a WSN. As explained in [10], under the assumption of spatially correlated sensor readings, the optimal linear approximation of $\mathbf{X}$ by $\mathbf{Y}$ such that $\|\hat{\mathbf{X}} - \mathbf{X}\|_F^2$ (Frobenius norm) is minimized will be

$$\hat{\mathbf{X}} = \mathbf{Y}\mathbf{Y}^{\dagger}\mathbf{X} = \mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{X} = \mathbf{Y}\mathbf{A} \tag{1}$$

where $\mathbf{Y}^{\dagger} = (\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T$ is the pseudo-inverse of the $\mathbf{Y}$ matrix; and $\mathbf{A} = \mathbf{Y}^{\dagger}\mathbf{X}$ is the *spatial correlation* between sensor data of the active nodes and the sleep nodes during $1 \leq t' \leq t$. Note that to estimate $\mathbf{A}$, one must also have the sleep mode data readings $\mathbf{X}$. We will denote this period a *training period* during which, the spatial correlation is learned from the sensor data with *all* sensor nodes activated.

Suppose one further assume that the spatial correlations $\mathbf{A}$ is slowly time-varying over a *testing period* following the training period, then one may use the $\mathbf{A}$ matrix computed during the training period over the testing period to predict the data that would be collected at the sleep nodes:

$$\hat{\mathbf{X}}_{test} \approx \mathbf{Y}_{test}\mathbf{A} \tag{2}$$

Assume the sensor nodes will be partitioned into $S$ disjoint subsets with each subset scheduled to a separate shift. One may generalize above mentioned training-testing operations as follows: Before starting multi-shift rotations, a common training period during which all sensor nodes are active will be carried out. Denoting $\mathbf{X}_i$ and $\mathbf{Y}_i$ to be the $i^{th}$ shift data of the sleep nodes,and active nodes respectively, one may define a data-centric performance criterion for $S$-shift partition as follows:

$$e = \max_{1 \leq i \leq S} e_i \tag{3}$$

where

$$e_i = \frac{\|\hat{\mathbf{X}}_i - \mathbf{X}_i\|^2}{\|\mathbf{X}_i\|^2} \tag{4}$$

Note that instead of absolute approximation error, a normalized error is used to deal with variations among different partitions. With above definitions, we say sensor nodes in the WSN are $(1 - e) \times 100\%$ data-covered by $s^{th}$-shift active nodes.

Once the $S$ partition is obtained, the corresponding spatial correlations $\mathbf{A}_i = \mathbf{Y}_i^{\dagger}\mathbf{X}_i$ may be computed to recover the data at the sleep nodes during $i^{th}$ shift (testing period).

### 2.2. Connectivity

The entire WSN thus may be described as a graph $G = \{V, E\}$ where $V$ is the set of indices of sensor nodes and $E$ is the set of connected edges among sensor nodes.

In a wireless network, two adjacent nodes $v$ and $v'$ are connected if the distance between them $d(v, v')$ is less than the wireless transmission range $r_d$. An edge $e$ then will be established between $v$ and $v'$ in $G$ to indicate that these two nodes are connected. Path is a sequence of nodes from each of its nodes there is an edge to the next node in the sequence. Denote $V_s$ to be the indices of an arbitrary subset of sensor nodes. If for any two nodes $v, v' \in V_s$, they are connected to each other by paths comprised by nodes in this subset, then the set of the nodes in $V_s$ form a *connected component* and will be a valid sensor sub-network.

In this work, for each subset of sensor nodes assigned to a particular shift, the connectivity of this subset will be examined using a Breadth First Search algorithm.

## 3. PROBLEM FORMULATION

### Multi-Shift Node Scheduling (MSNS) Problem

With above notations, the WSN Multi-shift node scheduling (MSNS) problem may be formulated as follows:
*Given a densely deployed WSN with graph $G = \{V, E\}$, partition G into S disjoint sub-graphs $\{G_s; 1 \leq s \leq S\}$ such that the maximum of data coverage error e is minimized subject to the constraint that each $G_s$ remains a connected component.*

Given any network with $N$ nodes, there are $S^{N-1}$ possible partition types. The computation complexity of exhaustive search grows exponentially with respect to $N$. For each partition type, the connectivity constraints and data coverage error can be checked in polynomial time. The problem is a NP-hard problem.

## 4. MULTI-SHIFT PARTITION ALGORITHM

In this work, the multi-shift partition algorithm which has polynomial time complexity is proposed. The algorithm leverages the hierarchical clustering and benefits from the assumption that sensors with high correlated data should be partitioned into different shift. In order to cluster, the distance between clusters with single node component is defined as data recovery error, which indicates the correlation between data from two nodes.

Let $X_k$ be the $k^{th}$ column of the matrix $\mathbf{X}$, and correspond to the data collected by the $k^{th}$ sensor node during the training period. The data correlation between node $k$ and node $j$ is

$$C_{kj} = \frac{X_k^T X_j}{\|X_k\|_2 \cdot \|X_j\|_2} \tag{5}$$

Alternatively, the data recovery error using node $k$'s data to estimate node $j$'s data or vice versa is represnted as:

$$d(k,j) = \|X_k - \hat{X}_k\|^2 / \|X_k\|^2 = 1 - C_{kj}^2 \tag{6}$$

where $\hat{X}_k = X_j X_j^{\dagger} X_k$ is the projection of $X_k$ to $X_j$ And For clusters with multiple components such as clusters $K, L$, the similarity distance between them is defined as

$$d(K,L) = \min_{k \in K, l \in L} d(k,l). \tag{7}$$

.

**Multi-Shift Node Scheduling Algorithm**

*Initialization*: Given data matrix Y, a $t \times n$ matrix with data columns from $n$ nodes, graph adjacency matrix $G$. Each node is a cluster.

*Clustering highly correlated nodes:*

1. Using Y, calculate the distance between any two clusters. Find two clusters $I, J$ with minimum distance $d(I, J)$. If sum of number of components in two clusters is less than or equal to S, merge them into a cluster.

2. Check all the clusters whether number of member is equal to S. If it is, all the components in this cluster quit from clustering.

3. Go back to 1 until no new merge can be conducted or all nodes quit from clustering.

*Partitioning highly correlation nodes:*

1. For each cluster, a member is randomly assigned into a shift. None of component nodes in same cluster can be assigned into the same shift.

2. Form the graph adjacency matrix $G_s$ for each shift. Check the connectivity of each shift and if all shifts form connected subgraphs, continue. if any of those shifts are not connected, go to 1.

*Local optimal feasible solution search:*

1. Nodes are allowed to be moved among shifts one by one. For each possible movement, we choose the one which minimizes the maximum of the data coverage errors $e_s$ and keeps subnetworks of all shifts connected. Stop when no move can be done or the move results in higher data coverage error.

## 5. EVALUATION AND EXPERIMENT

A real-world data set released by Intel-Berkeley Lab is used. These data samples are temperature readings taken from 54 Mica2Dot nodes over a period of 36 days with a sampling interval of 30 seconds. Occasional missing data are interpolated from the before and after samples from the same sensor nodes. Then, each data stream is sub-sampled at an interval of 1000 seconds. Several data records contain long segments of missing data and hence are excluded from this experiment. In the end, data taken from 49 sensors are used in this experiment. To learn the spatial correlation among sensor nodes, the first 27 hours of sensor records are used as training data.

It is assumed that the wireless communication range of each sensor node is 12 meters. Experiments are conducted by using the multi-shift partitioning algorithm. In 3 shift partitioning, we run the experiment 100 times. In each experiment, the three shifts remain connected. The data coverage errors of all 100 tries range from 0.45% to 0.55%. In other words, one achieves 3-fold sensor operating life-span extension while maintaining 99.5% data-recovery rate. The sensor location for different shift of one experiment is shown in Figure 1.

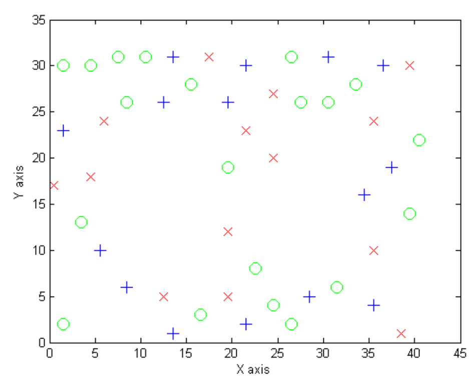


**Fig. 1**. Locations of Different Shift for 3 shift partition

Figure 2 shows the average distance of sensor with its closest neighbor as the function of number of partitions (shifts). As expected, when the number of shift increases, the minimum communication radius will also increase to maintain connectivity. In other words, even the spatial correlation among sensor nodes may support higher number of partitions, the energy consumption on maintaining connectivity will limit the number of shifts used.

In Figure 3, the achievable data recovery error versus number of shifts are plotted. As expected, as the number of shifts increases, the recovery error increases as well.

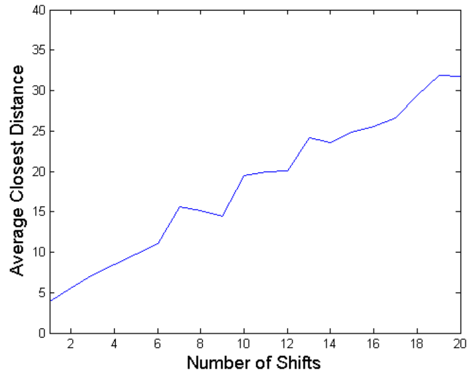


**Fig. 2**. average distance of sensor with its closest neighbor change vs the number of shifts

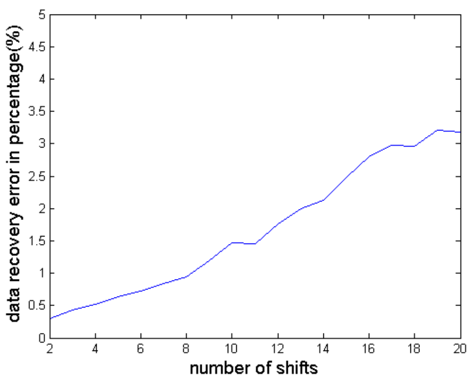


**Fig. 3**. data coverage error vs number of shifts

## 6. CONCLUSION

Energy is very limited in wireless sensor networks since usually each sensor is only equipped with a small battery. One way to extend the network lifespan is to keep only a part of sensors active while putting the others in sleep mode. To balance energy consumption, we propose a Multi-Shift Partition algorithm to schedule multiple sensor shifts. Our scheme can extend the network lifespan dramatically. Meanwhile, in order to maintain performance, the spatial correlation among sensors is exploited to recover data for sleep sensors and the data coverage error is defined as a criterion for performance of partitioning. Experiments show that the algorithm achieves a satisfying result over a real-world sensor dataset.

## 7. REFERENCES

[1] Y. Sankarasubramaniam I.F. Akyildiz, W. Su and E. Cayirci, "Wireless sensor networks: A survey," *Computer Networks*, pp. 393–422, 2002.

[2] T. H. Lai S. Kumar and J. Balogh, "On k-coverage in a mostly sleeping sensor network.," in *Proc. Mobicom '04*, pp. 144–158.

[3] Y. Zhang C. Lu R. Pless X. Wang, G. Xing and C. Gill, "Integrated coverage and connectivity configuration in wireless sensor networks," in *Proc. Int'l Conf. Embedded networked sensor systems*. ACM, pp. 28–39.

[4] H. Zhang and J. C. Hou, "Maintaining sensing coverage and connectivity in large sensor networks," in *Proc. Int'l Workship Theoretical and Algorithmic Aspects of Sensor, Ad Hoc Wireless, and Peer-to-Peer Networks*, 2004.

[5] A. Cerpa and D. Estrin, "Ascent: Adaptive self-configuring sensor networks topologies," in *Proc. INFOCOM*, 2002.

[6] J. Cheng S. Lu F. Ye, G. Zhong and L. Zhang, "Peas: A robust energy conserving protocol for long-lived sensor networks," in *Proc. ICDCS '03*, 2003, p. 2837.

[7] E. Bulut and I. Korpeoglu, "Dssp: A dynamic sleep scheduling protocol for prolonging the lifetime of wireless sensor networks," in *Proceeding of IEEE AINA*, 2007.

[8] P.G.Liaskovitis and C. Schurgers, "Leveraging redundancy in sampling-interpolation applications for sensor networks: A spectral approach," *ACM Trans. Sensor Networks*, vol. 7, pp. 12:1–12:28, September 2010.

[9] S. Slijepcevic and M. Potkonjak, "Power efficient organization of wireless sensor networks," in *Proc. ICC'01*, 2001, pp. 472–476.

[10] W. Liu X. Xu, Y. H. Hu and J. Bi, "Using data recovery models for multi-shift scheduling in wireless sensor networks.," in *ISCSCT (2)'08*. IEEE, pp. 777–782.

[11] D. Marsh R. Tynan, G. OHare and D. OKane, "Interpolation for wireless sensor network coverage," in *Proceeding of 2nd IEEE Workshop Embedded Networked Sensors*, pp. 123–131.