<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper presented at *IEEE International Conference on Communications, ICC 2015, London, United Kingdom, 8 June 2015 through 12 June 2015*.

Citation for the original published paper:

Du, R., Gkatzikis, L., Fischione, C., Xiao, M. (2015)
Energy efficient monitoring of water distribution networks via compressive sensing
In: *2015 IEEE International Conference on Communications (ICC)* (pp. 6681-6686).
IEEE conference proceedings
IEEE International Conference on Communications
<https://doi.org/10.1109/ICC.2015.7249390>

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-165840>

Energy Efficient Monitoring of Water Distribution Networks via Compressive Sensing

Rong Du*, Lazaros Gkatzikis[‡], Carlo Fischione*, Ming Xiao[†]

*Automatic Control Department, [†] Communication Theory Department
KTH Royal Institute of Technology, Stockholm, Sweden

[‡] Mathematical and Algorithmic Sciences Lab, France Research Center
Huawei Technologies Co. Ltd., Paris, France

Email: {rongd,carlofi,mingx}@kth.se, lazaros.gkatzikis@huawei.com

Abstract—The recent development of low cost wireless sensors enables water monitoring through dense wireless sensor networks (WSN). Sensor nodes are battery powered devices, and hence their limited energy resources have to be optimally managed. The latest advancements in compressive sensing (CS) provide ample promise to increase WSNs lifetime by limiting the amount of measurements that have to be collected. Additional energy savings can be achieved through CS-based scheduling schemes that activate only a limited number of sensors to sense and transmit their measurements, whereas the rest are turned off. The ultimate objective is to maximize network lifetime without sacrificing network connectivity and monitoring performance. This problem can be approximated by an energy balancing approach that consists of multiple simpler subproblems, each of which corresponds to a specific time period. Then, the sensors that should be activated within a given period can be optimally derived through dynamic programming. The complexity of the proposed CS-based scheduling scheme is characterized and numerical evaluation reveals that it achieves comparable monitoring performance by activating only a fraction of the sensors.

I. INTRODUCTION

Drinking water is an extremely valuable but limited resource. Thus, accurate and timely monitoring of water distribution networks has become a top priority in the context of smart cities. The major issues in water distribution networks are biological/chemical contamination and pipeline leaks, which result in huge waste of drinking water. To address those problems, wireless sensor networks (WSNs) have been deployed to monitor system parameters, such as pressure or water quality in Singapore [1] and Boston [2].

In contrast to existing works [3], where a small number of sensors has to be placed in key locations, here we consider a dense WSN system specifically designed for water distribution monitoring. Dense networks provide significant benefits. First, redundant nodes increase robustness to sensor failures and increase network lifetime [4]. In addition, dense WSNs are characterized by improved connectivity and enable more fine-grained monitoring.

Water monitoring sensors are generally deployed underground and hence replacing their battery or redeploying part of the WSN entails high costs. Thus, a major objective is to

maximize WSN lifetime such that maintenance cost is minimized. Densely deployed WSNs exploit the spatial correlation of sensed data so that only a subset of sensors has to be activated to sense and upload data through device-to-device communications. Thus, most of the sensors can be turned off or remain idle, whereas the rest guarantee sufficiently good monitoring performance.

In this paper, we address the inherent tradeoff between sensing performance and extended network lifetime in dense WSNs. Generally, a larger number of activated sensors improves monitoring performance but also increases the total amount of energy consumed and consequently network lifetime is reduced. However, measurements of densely deployed sensors exhibit in general strong spatial correlation and hence they are sparse in some basis, e.g., in Discrete Cosine Transform (DCT) domain [5]. In this context, compressive sensing (CS) dictates that the original signal can be recovered from a small number of measurements [6], [7]. Motivated by this observation, we devise a CS-based activation scheme that minimizes the overall energy consumption (sensing plus communication) towards maximizing network lifetime, without deteriorating monitoring performance. In addition, since the collected data should eventually reach a sink node in a multi-hop manner, we ensure that the resulting network of activated sensors is connected.

The key contributions of this paper are the following

- We propose the use of dense WSNs for water distribution monitoring, which facilitates the development of CS-based sensor activation and data gathering schemes that extend network lifetime.
- We devise a polynomial complexity activation algorithm that balances residual energy of sensors and eventually increases network lifetime.
- We evaluate through simulations the performance of the proposed scheme and compare it with an upper bound of network lifetime and greedy activation schemes. Our results indicate that the proposed algorithm extends lifetime significantly and approaches the upper bound if the network is dense enough.

The organization of the rest of the paper is as follows. An overview of related work is provided in Section II. We present the system model and formulate the problem of maximizing

This work is supported by the Wireless@KTH Seed Project LTE-based Water Monitoring Networks and it was conducted while L. Gkatzikis was a research associate at KTH Royal Institute of Technology.

Fig. 1. Data gathering in (a) Compressive Data Gathering (CDG); (b) Compressed Sparse Function (CSF); (c) Proposed scheme, where ϕ_i is a vector of size $M \ll N$.

lifetime in Section III. In Section IV, we devise a dynamic-programming approach to approximate maximum network lifetime and we analyze its complexity. The performance of the proposed algorithm is evaluated in Section V, and Section VI concludes our work.

II. RELATED WORK

A. WSNs for water monitoring

A relevant problem in monitoring water distribution network is that of optimal placement of a limited number of sensors. In this context, different objectives have been considered, e.g., minimizing the population exposed to a contaminant, the time between attack, and false alarm/misdetetection rate [8]. However, in case of abundant nodes, maximizing network lifetime is the most relevant problem.

Lifetime maximization has been extensively studied in WSNs. Seminal work [9] considers a static WSN where energy consumption is modeled as a function of the traffic flow routing decisions. In this case, the problem can be cast as a linear optimization problem, and the optimal amount of flow allocated to each link can be derived. In [10] the authors assume that each sensor may either transmit data to its one hop neighbors with unit energy consumption, or directly to the sink node through long range communication but at higher energy cost. They show that the problem of maximizing network lifetime is equivalent to the problems of flow maximization and energy-balancing. A similar topology is considered in [11], where network lifetime is pursued by balancing energy consumption in the network. However, these flow-based approaches cannot be directly applied to our case, where a cardinality constraint is imposed by CS. Thus, only the idea of energy balancing is adopted in our work, as shown later in detail.

B. Compressive sensing for data gathering

Energy savings in WSNs can be achieved by minimizing the amount of communicated data. Recently, compressive sensing (CS) has been proposed to enable reconstruction of signals from a limited number of measurements. Consider a network of N sensor nodes that have to transmit their sensing data denoted by vector $\mathbf{d} = [d_1, d_2, \dots, d_N]^T$ to the sink. Luo et al. proposed in [12] the Compressive Data Gathering (CDG) algorithm for data compression in the sensor side and data recovery at the sink node. Each sensor transmits only $M \ll N$ linear combinations of his own measurement and the received ones to its next hop node, as depicted in Fig. 1 (a).

Fig. 2. Architecture of a WSN for water distribution monitoring, sink node is connected to small LTE base station for long range communication

Inspired by CDG, Xu et al. proposed in [7] the Compressed Sparse Function (CSF) algorithm for data gathering that uses DCT to derive a sparse representation of the measured quantity. If the sink node collects M out of N data measurements of $\mathbf{d}$, it can use CS to estimate the remaining $N - M$ measurements of the sensors. The data transmission schedule of CSF is depicted in Fig. 1 (b). The grey nodes correspond to the ones that need to upload their measurements, while the white ones are relays. Here, we propose the scheme depicted in Fig. 1 (c) where only the activated nodes participate in both sensing and communication in CDG way, whereas the rest can be turned-off to save energy.

III. SYSTEM MODEL AND PROBLEM FORMULATION

We consider a water distribution network that is monitored by two tiers of nodes as shown in Fig. 2. The first tier consists of sensor nodes that are battery-powered and densely deployed in pipelines, whereas the second tier consists of sink nodes, which are grid-powered and deployed at the junctions of pipelines. This hierarchical architecture is typical in water distribution WSNs [2], [13], since it exhibits good scalability properties and addresses the main challenges of underground communications.

Sensor nodes are energy limited devices that are responsible for sensing, simple data processing and data-relaying to the sink nodes. Since the transmission range of sensor nodes is relatively small, compared to the length of a pipeline, for most of the sensors no sink node is in range. Thus, a multi-hop communication path has to be formed, so that sensed data reach a sink node. Due to the special signal propagation environment underground, it is natural to assume that each sensor can only transmit its data to nodes located in the same pipeline [14].

Sink nodes are responsible for network maintenance, data collection, data storing and transmission of collected data to a remote monitoring center. Since they are powered by the grid, their lifetime can be considered unlimited. They are also equipped with additional transceivers, e.g. LTE, that support long range communications. Thus, collected data can be directly forwarded to the monitoring center. A sink node is deployed in every junction to ensure network connectivity.

In this context, we consider the problem of maximizing network lifetime by scheduling activation of sensors. Since we use a CS-based data gathering scheme for monitoring, the network is said to be “dead” once the remaining nodes either cannot guarantee WSN monitoring performance or cannot reach a sink node (i.e., the network becomes disconnected). Since

communication between nodes located at different pipelines is impossible, we may consider the activation problem separately in each pipeline. Let s_l and s_r denote the leftmost and rightmost sink node respectively and $v_1, \dots, v_N$ the sensor nodes located in the pipeline. Without loss of generality, we assume that sensors are indexed according to their distance from s_l , i.e., v_1 is the closest one and v_N the most distant one.

We represent the sensor network within a pipeline as a communication graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. $\mathcal{V}$ is the set of vertices and consists of the 2 sink nodes at the two ends of the pipeline and N sensors. Edge set $\mathcal{E}$ captures the links among nodes that can communicate with each other. If r_i is the transmission range of node v_i and $d(v_i, v_j)$ is the distance between v_i and v_j , then $\forall v_i, v_j \in \mathcal{V}$, $\langle v_i, v_j \rangle \in \mathcal{E}$ iff $d(v_i, v_j) \leq r_i$. Although communication range is a function of transmission power, here we define r_i as the distance at which communication can be considered energy efficient. Generally, in high path loss environments significant energy savings can be achieved by adopting multihop over single hop communications. Thus, in a dense network, we may assume that r_i is the range resulting from minimum transmission power [15].

Let $\mathcal{N}(v_i) = \{v_j | \langle v_i, v_j \rangle \in \mathcal{E}\}$ be the set of neighbors of node v_i . The upstream set (UNS) of v_i is $\mathcal{N}_+(v_i) = \{v_j | v_j \in \mathcal{N}(v_i) \wedge (j < i)\}$ and $\mathcal{N}_-(v_i) = \{v_j | v_j \in \mathcal{N}(v_i) \wedge (j > i)\}$ is the downstream set (DNS) of v_i . We also denote by $\mathcal{N}'_-(v_i) = \mathcal{N}_-(v_i) \setminus \{s_r\}$ the set of nodes in the DNS of v_i excluding sink node s_r .

Time is slotted. In each timeslot t a sensor can be either activated to sense and transmit data, or set to idle mode to save energy. For simplicity, we normalize overall energy consumption of a sensor (including operation, sensing, receiving and transmitting data) to 1 energy unit per slot if it is activated and 0 otherwise. This is a reasonable assumption since the overall energy consumption of a sensor is mainly determined by its state and not the exact amount of data that it transmits/receives within a timeslot [16]. Sensor node v_i has an initial energy budget of b_i , while sink nodes have unlimited energy.

To meet the specific application requirements, the set of activated sensors in each timeslot should always guarantee connectivity and monitoring quality. Let binary variable $x_i(t)$ indicate whether sensor i is activated in timeslot t and $\mathbf{x}(t) = [x_1(t), \dots, x_N(t)]^T$ be the corresponding activation schedule of the WSN. The set of activated sensors together with the two sink nodes are denoted by $\mathcal{V}_A(t)$, and the corresponding communication graph by $G(\mathcal{V}_A(t))$ has to be always connected. We assume that a scheme based on CSF [7] is used for data gathering (see Section IV), and compressive sensing is used for data estimation. Then, monitoring quality requirement is met if the number of measurements M exceeds a threshold, namely if

$$M \geq ck \log N \triangleq M_{\min}, \quad (1)$$

where c is a positive constant and k is the sparsity of data [6]. Thus, monitoring quality requirement can be represented as a cardinality constraint on $\mathcal{V}_A(t)$.

Maximum network lifetime T is the number of timeslots until either the connectivity or cardinality constraint cannot be

satisfied; hence, we may formulate the problem of maximum lifetime as

$$\max_{\mathbf{x}(t), t=1, \dots, T} T \quad (2a)$$

$$\text{s.t.} \quad \sum_{v_i \in \mathcal{V}_A(t)} x_i(t) \geq M_{\min}, \quad \forall t \quad (2b)$$

$$G(\mathcal{V}_A(t)) \text{ is connected}, \quad \forall t \quad (2c)$$

$$\sum_{t=1}^T x_i(t) \leq b_i, \quad \forall i \in \mathcal{V} \quad (2d)$$

$$x_i(t) \in \{0, 1\}, \quad \forall i \in \mathcal{V}, \quad \forall t, \quad (2e)$$

where constraints (2b) and (2c) ensure that the network satisfies both cardinality and connectivity constraint in each timeslot, and constraint (2d) captures the limited energy budget of sensor nodes. By including the two sink nodes in connectivity constraint (2c), we impose that the activated sensors will be uniformly distributed in the pipeline, so that the whole pipeline is monitored.

Problem (2) is particularly challenging due to the binary nature of the variables, the non-convex connectivity constraint, and the additional cardinality constraint imposed by compressive sensing. If the latter is relaxed, Problem (2) can be solved as a maximum flow instance, similarly to [9], [10]. However, the additional cardinality constraint makes the problem of special interest, and the proposed solution approach constitutes one of the core contributions of this paper.

An upper bound of network lifetime, $\bar{T}$, can be easily derived: $\bar{T} = \sum_{i=1}^N b_i / M_{\min} \geq T$. Notice that this upper bound corresponds to a network instance, where the total energy is perfectly balanced among a set of nodes of cardinality $M_{\min}$ such that compressive sensing requirement (1) is met.

It has been shown that in specific network scenarios, the problems of lifetime maximization, maximum flow and energy balancing are equivalent [10], [17]. This observation motivates us to approximate the optimal solution of Problem (2) by balancing residual energy of sensors in each time slot. A side benefit of energy balancing is that it enables a more efficient handling of node failures.

The main idea is to approximate the lifetime maximization problem through a series of online subproblems. In each timeslot: 1) First, we minimize the number of sensors m that should be activated, in an attempt to minimize the total energy consumption in each timeslot. The exact value of m is determined by the cardinality and connectivity constraints. 2) Given m , find which m connected sensors should be activated to balance the residual energy. The latter can be formulated as:

$$\max_{\mathbf{x}} \sum_{i \in \mathcal{V}} x_i p_i \quad (3a)$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{V}_A} x_i = m \quad (3b)$$

$$G(\mathcal{V}_A) \text{ is connected} \quad (3c)$$

$$x_i \in \{0, 1\} \quad \forall i \in \mathcal{V}, \quad (3d)$$

where p_i is the residual energy of node v_i , m is the minimum number of sensors that satisfies both cardinality and connectivity constraints, and for simplicity we have discarded time

Algorithm 1 GBS Algorithm

Input: The adjacency matrix of nodes $\mathbf{A}$.**Output:** Minimum number of sensors for connectivity.

```

1: Set  $M \leftarrow 0, k \leftarrow 0, v_0 \leftarrow s_l$ .
2: while  $s_r \notin \mathcal{N}_-(v_k)$  do
3:   if  $\mathcal{N}_-(v_k) \neq \emptyset$  then
4:      $k \leftarrow \max\{i : v_i \in \mathcal{N}_-(v_k)\}, M \leftarrow M + 1$ 
5:   else
6:      $M \leftarrow +\infty$ . Break
7:   end if
8: end while
9: return  $M$ 

```

indices. Notice that once the battery of sensor i is depleted, $p_i = 0$, it should be removed from set $\mathcal{V}$. Sink nodes solve this problem online in each timeslot to determine which sensor nodes should be activated according to their residual energy.

Finding the minimum number of sensors m is easy. First, we find the minimum number of sensors M that satisfies connectivity constraint, e.g. by finding the shortest path from s_l to s_r . Then, we have $m = \max\{M, M_{\min}\}$. Given m , Problem (3) is a cardinality-constrained Maximum Weight Connected Subgraph (MWCS) problem, which is NP-hard [18]. Nevertheless, under certain assumptions that generally hold in the considered scenario, we demonstrate that it can be solved efficiently.

IV. AN OPTIMAL SENSOR ACTIVATION SCHEDULE TO BALANCE RESIDUAL ENERGY

In this section, we derive an efficient online algorithm for the solution of problem (3), and characterize its complexity. Based on the proposed algorithm, we devise an easy-to-implement network protocol.

A. Optimal sensor activation in a pipeline

We introduce the following assumption, which generally holds for WSNs deployed in water distribution pipelines.

Assumption 1: All sensor nodes are characterized by the same communication range $r_i = r$ and are deployed on a line.

Under this assumption, we derive a solution algorithm for the two subproblems.

Calculating the minimum number of nodes for connectivity

This subproblem can be solved efficiently through the Greedy Based Search (GBS) Algorithm that we present in Algorithm 1. Starting from sink s_l , in each step we calculate the rightmost node belonging to its neighbor set, which then becomes the starting node for the next iteration. If in any iteration the DNS of current node is empty (line 3), then the network is disconnected, no feasible solution can be found and network lifetime has been reached.

Theorem 1: Consider optimization problem (3). Under Assumption 1, Algorithm 1 calculates the minimum number of sensors that have to be activated to guarantee connectivity.

Proof: Due to limited space, the proof is omitted. It can be found in technical report¹. ■

Algorithm 2 SAE Algorithm

Input: $\mathbf{A}$, $\mathbf{p}$ and the number of sensors m to be activated.**Output:** The optimal set of sensor nodes $\mathcal{V}_A$ to be activated.

```

1: Construct a matrix  $\mathbf{G} = \{g_{ij}\}$  of size  $(N+1) \times (m+1)$ , where all its elements are initially set to  $-\infty$ .
2: Construct a node matrix  $\mathbf{H} = \{h_{ij}\}$  of size  $(N+1) \times (m+1)$  where all its elements are  $-1$  initially.
3: for all  $v_k : s_r \in \mathcal{N}_-(v_k)$  do
4:    $g(v_k, 0) = 0$  and set  $g_{k1} = g(v_k, 0)$ 
5:    $h_{k1} = N + 1$ 
6: end for
7: for  $i = 2$  to  $m$  do
8:   for  $k = N$  to  $i$  do
9:      $g_{ki} \leftarrow \max_{j: v_k \in \mathcal{N}_-(v_j)} \{g_{j(i-1)} + p_k\}$ 
10:     $h_{ki} \leftarrow \operatorname{argmax}_{j: v_k \in \mathcal{N}_-(v_j)} \{g_{j(i-1)} + p_k\}$ 
11:   end for
12: end for
13:  $g_{1(m+1)} \leftarrow \max_{v_j \in \mathcal{N}_-(s_l)} \{g_{jm}\}, h_{1(m+1)} \leftarrow \operatorname{argmax}_{j: v_j \in \mathcal{N}_-(s_l)} \{g_{jm}\}$ 
14:  $c \leftarrow 1, \mathcal{V}_A \leftarrow \emptyset$ .
15: for  $k = M + 1$  to  $2$  do
16:    $\mathcal{V}_A = \mathcal{V}_A \cup \{h_{ck}\}, k \leftarrow h_{ck}$ 
17: end for
18: return  $\mathcal{V}_A$ 

```

Algorithm 3 Optimal activation schedule for problem (3)

Input: Adjacency matrix $\mathbf{A}$, and remaining battery $\mathbf{p}$.**Output:** A set of sensor nodes $\mathcal{V}_A$ that need to be activated.

```

1:  $M \leftarrow \text{GBS}(\mathbf{A})$ 
2: if  $M < +\infty$  then
3:   Calculate  $M_{\min} = ck \log N$ 
4:    $\mathcal{V}_A \leftarrow \text{SAE}(\mathbf{A}, \mathbf{p}, \max\{M, M_{\min}\})$ 
5:   return  $\mathcal{V}_A$ 
6: else
7:   return  $\emptyset$ 
8: end if

```

Deriving the maximum weight connected subset of sensors

Given the number of sensors that have to be activated, we need to determine which ones should be selected. Under Assumption 1, we may derive a dynamic programming approach to select which ones to activate. Assume that in a given state, v_i has been selected and k additional sensors have to be activated out of $\{v_{i+1}, \dots, v_N\}$ such that a connected subgraph of maximum sum weight (total residual energy) is formed. Let $g(v_i, k)$ denote the maximum total residual energy out of all the possible subsets of k activated sensors. Then, we have

$$g(v_i, k) = \begin{cases} \max_{v_j \in \mathcal{N}'_-(v_i)} \{g(v_j, k-1) + p_j\} & \text{if } \mathcal{N}'_-(v_i) \neq \emptyset \\ -\infty & \text{otherwise} \end{cases} \quad (4)$$

For $k = 1$, the selected sensor should be a neighbor of s_r . Consequently, for any sensor $v_i \in \mathcal{V}$ we have

$$g(v_i, 0) = \begin{cases} 0 & \text{if } s_r \in \mathcal{N}_-(v_i) \\ -\infty & \text{otherwise} \end{cases} \quad (5)$$

Based on recursive function $g(\cdot)$, we devise the Sensor Activation on Edge (SAE) Algorithm. The exact steps are described in Algorithm 2, where $\mathbf{A}_{(N+2) \times (N+2)}$ is the adjacency matrix

¹<https://sites.google.com/site/rongdusjtu/ProofsInICC15Energy.pdf>

of the WSN and $\mathbf{p} = [p_1, \dots, p_N]^T$ is the residual energy of sensors. Then, the following theorem can be derived.

Theorem 2: Consider optimization problem (3). Under Assumption 1 and if m is the minimum number of sensors that satisfies both connectivity and cardinality constraints, then Algorithm 2 calculates the optimal activation schedule.

Proof: The proof can be found in technical report¹. ■

SAE and GBS algorithms are the building blocks of Algorithm 3 that provides the optimal solution to Problem (3) under Assumption 1. The complexity of Algorithm 3 is given in the following proposition.

Proposition 1: Let ρ be the density of the deployed sensor network, N the number of sensors, and r the communication range of each sensor. Then, the time and space complexity of Algorithm 3 are $O(\max\{N^2, kNr\rho \log N\})$ and $O(N^2)$ respectively.

Proof: The proof can be found in technical report¹. ■

According to Proposition 1, the proposed Algorithm 3 is of low complexity and hence can be executed at the sinks to derive the set of sensors to activate in each timeslot.

B. A protocol for optimal sensor activation

Next, we describe how the proposed Algorithm 3 can be translated into a protocol and applied to a real WSN for water monitoring.

1) *Data Gathering:* For each timeslot, the sensors that have to be activated are decided by sinks according to Algorithm 3. The rest of the nodes switch into sleep mode to save energy. The schedule is broadcasted by the sink. It is natural to assume that sinks are equipped with transceivers that support communication with all sensors. Otherwise the schedule could be forwarded in a multi-hop manner. In any case, such a coordination of nodes is supported by existing standards such as IEEE 802.15.4, e.g., [19]. Activated sensors sense and transmit their measurements to the sink node through multiple hops in CDG way. Notice that here we only devise the activation schedule, while the actual routing is beyond the scope of this work. A straightforward approach is to forward data towards the closest sink. Once the sinks receive the data, they update system parameters $\mathbf{p}$, $\mathbf{A}$.

2) *Data Recovery:* In each timeslot, the data of the idle sensors have to be recovered through compressive sensing. As the activated nodes transmit data in CDG way, the sink node needs to first recover the original data, and then estimate the data from the inactive nodes similar to [7]. Consider a setting of N sensors where the activated sensor nodes in this timeslot are labeled $v_{a_1}, v_{a_2}, \dots, v_{a_m}$. Let the estimated data of the m activated nodes be $\mathbf{d}_r = [d_{a_1}, d_{a_2}, \dots, d_{a_m}]^T$. Then, we have

$$\mathbf{d}_r = \begin{bmatrix} p_1(a_1) & p_2(a_1) & \dots & p_N(a_1) \\ p_1(a_2) & p_2(a_2) & \dots & p_N(a_2) \\ \vdots & \vdots & \ddots & \vdots \\ p_1(a_m) & p_2(a_m) & \dots & p_N(a_m) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \triangleq \mathbf{P}_A \mathbf{x}, \quad (6)$$

Fig. 3. Performance of Compressive Sensing with different m/N -ratio

where $\mathbf{P}_A$ is the corresponding submatrix of $\mathbf{P} = \{p_i(j)\}$, which is the Type-IV DCT function base and $p_i(j) = \sqrt{2/N} \cos[\pi(i-0.5)(j-0.5)/N]$. Then, we may use CS to estimate the coefficient $\mathbf{x}$ by solving

$$\hat{\mathbf{x}} = \arg \min \|\mathbf{x}\|_{l_1} \quad \text{s.t.} \quad \|\mathbf{d}_r - \mathbf{P}_A \mathbf{x}\|_{l_2} \leq \epsilon. \quad (7)$$

Given the estimate of coefficient $\hat{\mathbf{x}}$, we can estimate the data of the idle sensors through $\hat{d}_i = [p_1(i) \ p_2(i) \ \dots \ p_N(i)] \hat{\mathbf{x}}, \forall i \notin \{a_1, \dots, a_m\}$.

V. NUMERICAL EVALUATION

In this section, we evaluate the performance of the proposed Algorithm 3. First, we validate applicability of CS in water distribution networks through EPANET², a software that models water distribution piping systems. We generate a water distribution network and simulate water pressure over a pipeline equipped with $N = 50$ sensors. We randomly select m sensors to be activated. According to the IDs of the sensors, $\mathbf{P}_A$ is generated based on (6). The coefficient $\hat{\mathbf{x}}$ is derived by using orthogonal matching pursuit (OMP) Algorithm [20] to solve (7). Then, the estimation of original data $\hat{\mathbf{d}}$ is performed according to $\hat{\mathbf{x}}$. The recovery error is defined as $\|\hat{\mathbf{d}} - \mathbf{d}\|_2 / \|\mathbf{d}\|_2$ and is depicted for different values of m in Fig. 3. The recovery error is about 5% when m/N is 0.2, which can be considered small enough for most applications.

Next, in order to investigate network lifetime performance, we consider a simulation set up where the length of the pipeline is normalized to 1. Two sink nodes are deployed at the endpoints of the pipeline, namely at point 0 and point 1. N sensors are uniformly distributed in the pipeline in the range of $(0, 1)$. The initial energy budget of each sensor node is randomly set to be $100 + \omega$, where $\omega \sim \mathcal{N}(0, 1)$. The number of nodes are either 100 or 150. The energy consumed by an activated node in a timeslot is normalized to 1.

Two greedy algorithms are considered for comparison. The first one, called Greedy Based Search with Random activation (GBS+R), is based on Algorithm 1 to select M sensor nodes so as to ensure connectivity. If $M > M_{\min}$, these M nodes are selected for activation. Otherwise, additional $M_{\min} - M$ sensors are selected randomly. The second one, Greedy Based Search with Maximum remaining energy (GBS+M), also uses GBS Algorithm to find M sensor nodes that guarantee connectivity. Here, for the remaining $M_{\min} - M$ sensors, the ones of maximum residual energy are selected. It can be considered as

²<http://www.epa.gov/nrmrl/wswrd/dw/epanet.html>

Fig. 4. Comparison of the network's average lifetime, maximum lifetime and minimum lifetime for the three algorithms under different transmission range with (a) $N = 100$, $M_{\min} = 15$; (b) $N = 100$, $M_{\min} = 25$; (c) $N = 150$, $M_{\min} = 22$; (d) $N = 150$, $M_{\min} = 37$

a mixture of the GBS+R Algorithm and the SAE algorithm. The time complexity of GBS+R and GBS+M is $O(N)$ and $O(N \log N)$ respectively.

The resulting lifetime of the proposed algorithm and the greedy ones is depicted in Fig. 4 for a diverse set of scenarios. Average network lifetime using the proposed SAE Algorithm is significantly longer. When the ratio of transmission range to pipeline's length is small enough, the average lifetime of the network is very limited. The network is sparse and it is easy to become disconnected once a small number of vital nodes die. However, as the ratio increases from 0.08 to 0.20, average network lifetime increases significantly. It approaches the upper bound of network maximum lifetime when the ratio is large enough, for example, $r = 0.16$ for the case that $N = 100$, $M_{\min} = 15$.

By comparing the three algorithms, we see that SAE outperforms GBS+M which in turn is better than GBS+R. Our numerical results validate that it is reasonable to select the nodes of highest remaining battery to prolong network lifetime.

VI. CONCLUSION

In this paper, we considered the problem of optimal sensor activation in water distribution networks with the objective of maximizing network lifetime. We derived an approximation algorithm based on the notion of energy balancing. Under certain conditions, we showed that the latter can be optimally solved through dynamic programming. Our complexity analysis revealed that the proposed approach is applicable in real scenarios and can be executed at sink nodes. The simulation results revealed that the proposed algorithm outperforms greedy schemes by extending lifetime up to 20% and performs close to the upper bound, if the network is dense enough.

REFERENCES

- [1] A. J. Whittle, L. Girod, A. Preis, M. Allen, H. B. Lim, M. Iqbal, S. Srirangarajan, C. Fu, K. J. Wong, and D. Goldsmith, "Waterwise@sg: A testbed for continuous monitoring of the water distribution system in singapore," *Water Distribution System Analysis, WSDA*, 2010.
- [2] I. Stoianov, L. Nachman, S. Madden, and T. Tokmouline, "Pipenet: A wireless sensor network for pipeline monitoring," in *Proc. of the 6th International Symposium on Information Processing in Sensor Networks*. IEEE, 2007, pp. 264–273.
- [3] A. Bertoldi, G. Athanasiou, and C. Fischione, "Optimal sensor placement for bacteria detection in water distribution networks," in *Proc. of the 6th International Symposium on Communications, Control and Signal Processing*. IEEE, 2014, pp. 248–253.
- [4] A. Krause, R. Rajagopal, A. Gupta, and C. Guestrin, "optimization of sensor placements and balanced schedules," *IEEE Trans. on Automatic Control*, vol. 56, no. 10, pp. 2390–2405, 2011.
- [5] B. He, Y. Li, H. Huang, and H. Tang, "Spatial-temporal compression and recovery in a wireless sensor network in an underground tunnel environment," *Knowledge and Information Systems*, pp. 1–17, 2014.
- [6] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [7] L. Xu, X. Qi, Y. Wang, and T. Moscibroda, "Efficient data gathering using compressed sparse functions," in *Proc. of IEEE INFOCOM*. IEEE, 2013, pp. 310–314.
- [8] J.-P. Watson, H. J. Greenberg, and W. E. Hart, "A multiple-objective analysis of sensor placement optimization in water networks," *Proc. ASCE World Water Environ. Resources Congr.*, pp. 456–465, 2004.
- [9] J.-H. Chang and L. Tassiulas, "Maximum lifetime routing in wireless sensor networks," *IEEE/ACM Trans. on Networking (TON)*, vol. 12, no. 4, pp. 609–619, 2004.
- [10] A. Jarry, P. Leone, O. Powell, and J. Rolim, "An optimal data propagation algorithm for maximizing the lifespan of sensor networks," in *Distributed Computing in Sensor Systems*. Springer, 2006, pp. 405–421.
- [11] F. Ren, J. Zhang, T. He, C. Lin, and S. K. Ren, "Ebrp: energy-balanced routing protocol for data gathering in wireless sensor networks," *IEEE Trans. on Parallel and Distributed Systems*, vol. 22, no. 12, pp. 2108–2125, 2011.
- [12] C. Luo, F. Wu, J. Sun, and C. W. Chen, "Compressive data gathering for large-scale wireless sensor networks," in *Proc. of Mobicom*. ACM, 2009, pp. 145–156.
- [13] Z. Sun, P. Wang, M. C. Vuran, M. A. Al-Rodhaan, A. M. Al-Dhelaan, and I. F. Akyildiz, "Mise-pipe: Magnetic induction-based wireless sensor networks for underground pipeline monitoring," *Ad Hoc Networks*, vol. 9, no. 3, pp. 218–227, 2011.
- [14] S. Xiong, L. Yu, H. Shen, C. Wang, and W. Lu, "Efficient algorithms for sensor deployment and routing in sensor networks for network-structured environment monitoring," in *Proceedings of IEEE INFOCOM*. IEEE, 2012, pp. 1008–1016.
- [15] R. Min and A. Chandrakasan, "A framework for energy-scalable communication in high-density wireless networks," in *Proceedings of the International Symposium on Low Power Electronics and Design*.
- [16] L. Gkatzikis, G. Paschos, and I. Koutsopoulos, "Medium access games: The impact of energy constraints," in *5th International Conference on Network Games, Control and Optimization (NetGCoP)*, 2011.
- [17] H. Zhang and H. Shen, "Balancing energy consumption to maximize network lifetime in data-gathering sensor networks," *IEEE Trans. on Parallel and Distributed Systems*, vol. 20, no. 10, pp. 1526–1539, 2009.
- [18] E. Álvarez-Miranda, I. Ljubić, and P. Mutzel, "The maximum weight connected subgraph problem," in *Facets of Combinatorial Optimization*. Springer, 2013, pp. 245–270.
- [19] P. Park, C. Fischione, and K. H. Johansson, "Adaptive IEEE 802.15.4 protocol for energy efficient, reliable and timely communications," in *Proc. of the 9th ACM/IEEE international conference on information processing in sensor networks*. ACM, 2010, pp. 327–338.
- [20] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.