

## Research Article

# A Mutual Algorithm for Optimizing Distributed Source Coding in Wireless Sensor Networks

Nashat Abughalieh,<sup>1</sup> Kris Steenhaut,<sup>1,2</sup> Bart Lemmens,<sup>1,2</sup> and Ann Nowé<sup>1</sup>

<sup>1</sup>Department of Electronics and Informatics (ETRO), Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium

<sup>2</sup>Erasmushogeschool Brussel—IWT, Nijverheidskaai 170, 1070 Brussel, Belgium

Correspondence should be addressed to Nashat Abughalieh, nashatg@gmail.com

Received 15 July 2011; Revised 10 November 2011; Accepted 16 November 2011

Academic Editor: Yuhang Yang

Copyright © 2012 Nashat Abughalieh et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Wireless Sensor Networks (WSNs) are composed of small wireless nodes equipped with sensors, a processor, and a radio communication unit, all normally powered by batteries. For most WSN applications, the network is expected to function for several months or years. In the common monitoring application scenario, adjacent nodes in a WSN often sense spatially correlated data. Suppressing this correlation can significantly improve the lifetime of the network. The maximum possible network data compression can be achieved using distributed source coding (DSC) techniques when nodes encode at Slepian-Wolf rates. This paper presents contributions to the lifetime optimization problem of WSNs in the form of two algorithms: the Updated-CMAX (UCMAX) power-aware routing algorithm to optimize the routing tree and the Rate Optimization (RO) algorithm to optimize the encoding rates of the nodes. The two algorithms combined offer a solution that maximizes the lifetime of a WSN measuring spatially correlated data. Simulations show that our proposed approach may significantly extend the lifetime of multihop WSNs with nodes that are observing correlated data.

## 1. Introduction

Wireless Sensor Networks have a wide range of possible applications like environmental monitoring, home automation, military, industrial, and medical applications [1]. Network designers must consider factors such as the environment, cost, and hardware, while engineering a particular WSN. Different applications will prompt different architectural constraints and requirements [2]. For some applications, network designers will need to focus on bounded delivery time of sensed events to the Base Station (BS), for example, Tsunami warning systems. Other applications require the WSN to function for several months or years before being replaced, and designers are thus more concerned about the lifetime of the network, such as in environmental monitoring systems [3]. Since the network lifetime has been the main challenge in the design of many WSN applications [4], we address the problem of maximizing the network lifetime in this paper.

The nodes in WSNs are mostly powered by batteries. The energy of the batteries is utilized by the main building

blocks of a node: the sensors, the processor, and the radio unit. The radio is known to be the most energy consuming component of the node [5, 6]. For most WSNs, the radio unit has four functional modes: sleep, active, transmit and receive. The transmit, and receive modes have the highest power consumption while the sleep mode has the lowest power consumption. To improve the lifetime of the network (the lifetime of a WSN has many definitions [7]: some consider it to be the timespan from network startup to the death of a certain percentage of the nodes, others define it as the timespan from startup to the loss of coverage of a certain percentage of the monitored area), the node's radio has to be switched to sleep mode as much as possible. In this paper, we accomplish this by reducing the size of the packets through the application of lossless compression techniques which remove the redundancy in the spatially correlated sensed data. Lossless compression can be achieved by Source Coding techniques. Two lossless Source Coding methodologies for WSNs are described in the literature: Explicit Communications (EC) [8–10] and distributed source coding (DSC) [11–14].

The EC coding technique eliminates the redundancy in the spatially correlated sensed data while routing all data to the Base Station. Each node compresses its own sensed data according to the data flow passing through it from other nodes in the routing tree. The problem of finding the optimum routing path that achieves maximum compression with minimum network power consumption using EC is proven to be NP-hard [9]. The EC encoding process requires intensive processing at each sensor node to compress its sensed data, especially when the node has to forward compressed data from other nodes: the routing node needs to uncompress the data flow before being able to encode its own data in order to remove the redundancy due to spatial correlation.

The other methodology used for applying lossless source coding in WSNs is DSC. The concept of DSC was introduced by Slepian and Wolf in [15]. They derived the admissible rate region of two correlated sources and proved that two source encoders can compress their input data to a total rate which equals their joint entropy, without communication between the encoders, on condition that they are jointly decoded. Since, many authors have proposed source encoding systems that almost achieve the Slepian-Wolf theoretical limit. In [16], the authors used Coset Coding for compressing one of the two sources, while using the second source's data at the decoder to predict the data of the first source. In [17, 18], the authors used turbo codes and reached near the Slepian-Wolf theoretical limit. Their coding techniques are based on sending the data of the first node to the base station without coding, while encoding the second node's output with a turbo encoder and then send some parity bits of the encoder's output to the Base Station. The decoder at the BS uses the parity bits together with the data from the first node to estimate the second node's data. In [19–21], the authors implemented DSC using Low-Density Parity-Check (LDPC) codes. Turbo codes and LDPC codes enable DSC implementations which almost reach the Slepian-Wolf theoretical limit [17, 21]. Using the DSC theory, Cristescu et al. [22] studied how optimizing the rates of the nodes can minimize the network's total power consumption, but they did not address optimizing the lifetime of the network.

We propose contributions to the lifetime maximization problem of WSNs through the formulation of the problem's optimization equations and the development of algorithms which assign data rates and routing paths to the network nodes. The paper is organized as follows: we review the prior work on DSC for WSN in Section 2. We derive a system of equations for the optimal DSC rates and introduce the routing and rate assignment algorithms in Section 3. In Section 4, simulations results of the optimization algorithms are shown and discussed. We conclude the paper in Section 5.

## 2. Prior Work on DSC for WSNs

For a random source  $X$ , a rate  $R \geq H(X)$  is sufficient to transmit  $X$  over a reliable channel to the BS. If we have two *independent and identically distributed (i.i.d.)* sources  $(X_1, X_2)$ , as shown in Figure 1, and they are encoded

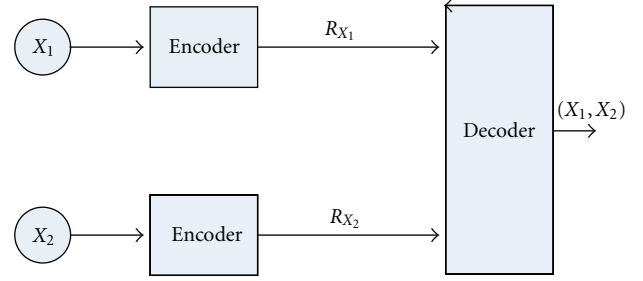


FIGURE 1: Slepian-Wolf coding for two correlated sources.

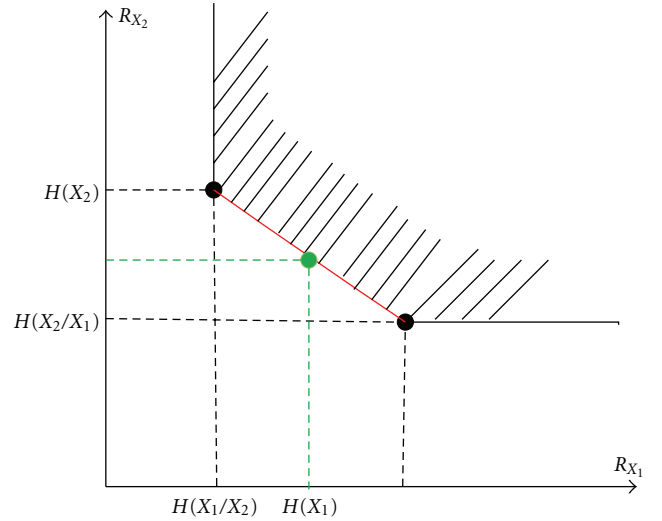


FIGURE 2: Rate region for Slepian-Wolf encoding.

separately, a total rate  $R = R_{X1} + R_{X2} \geq H(X1) + H(X2)$  is required. However, as Slepian and Wolf have shown in their seminal paper [15], when both sources are correlated, a total rate  $R = R_{X1} + R_{X2} \geq H(X1, X2)$  is sufficient, even when the two sources are encoded separately, as long as they are decoded jointly.

The achievable rate region according to the Slepian-Wolf coding theory is determined by the following equations:

$$\begin{aligned} R_{X1} &\geq H\left(\frac{X1}{X2}\right), \\ R_{X2} &\geq H\left(\frac{X2}{X1}\right), \end{aligned} \quad (1)$$

$$R_{X1} + R_{X2} \geq H(X1, X2).$$

The solution of this system of equations is shown in Figure 2. The minimum theoretical rate for the two correlated sources scenario is shown in Figure 2 with a red line. The two black dots at the corners of the optimum rate region correspond to the following rates:

$$\begin{aligned} R_{X1} &= H\left(\frac{X1}{X2}\right), & R_{X2} &= H(X2), \text{ or} \\ R_{X1} &= H(X1), & R_{X2} &= H\left(\frac{X2}{X1}\right). \end{aligned} \quad (2)$$

The Slepian-Wolf theory can be generalized to many sources [23]. If  $X_1, X_2, \dots, X_n$  are *i.i.d.*, but spatially correlated sources, then the set of rate vectors achievable using distributed source coding with separate encoders and joint decoder is defined by

$$R(S) \geq H\left(\frac{X(S)}{X(S^C)}\right) \quad (3)$$

for all  $S \subseteq \{1, 2, 3, \dots, n\}$ , where

$$R(S) = \sum_{i \in S} R_i. \quad (4)$$

Cristescu et al. [22] used the generalized Slepian-Wolf theory of multiple sources to optimize the encoding rates of WSN nodes. They set out to find an optimal transmission structure, that is, routing tree, and rate allocation for the nodes of a multihop WSN with multiple sensors  $X_i$ ,  $\{i = 1, 2, 3, \dots, n\}$  and one BS that minimizes a certain cost function which reflects the network's total power consumption:

$$\{R_i, d_i\}_{i=1}^N = \arg \min_{\{R_i, d_i\}_{i=1}^N} \sum_{i=1}^N F(R_i) \cdot d_i, \quad (5)$$

under the constraint of Slepian-Wolf encoding rates

$$\sum_{i \in S} R_i \geq H\left(\frac{X(S)}{X(S^C)}\right), \quad (6)$$

where  $R_i$  and  $d_i$  are the transmission rate of node  $X_i$  and the total weight (cost) associated to the routing links from node  $X_i$  to the Base Station (BS), respectively. They showed that for this specific type of cost function, the optimization problem can be decomposed into two separate problems: routing and rate optimization. Substituting the Shortest Path Tree (SPT) as the optimum routing tree, the optimization problem of (5) is reduced to

$$R_i = \arg \min_{R_i} \sum_{i=1}^N R_i d_{\text{SPT}}. \quad (7)$$

When numbering nodes in increasing order by weight of the routing paths from each node to the BS, so that nodes  $X_1, X_2, \dots, X_n$  have SPT weights  $d_{\text{SPT}}(X_1) \leq d_{\text{SPT}}(X_2) \leq \dots \leq d_{\text{SPT}}(X_N)$ , the solution to (7) under the constraint of Slepian-Wolf encoding is [22]

$$\begin{aligned} R_1 &\geq H(X_1), \\ R_2 &\geq H\left(\frac{X_2}{X_1}\right), \\ R_3 &\geq H\left(\frac{X_3}{X_1}, X_2\right) \\ &\vdots \\ R_N &\geq H\left(\frac{X_N}{X_{N-1}}, X_{N-2}, \dots, X\right). \end{aligned} \quad (8)$$

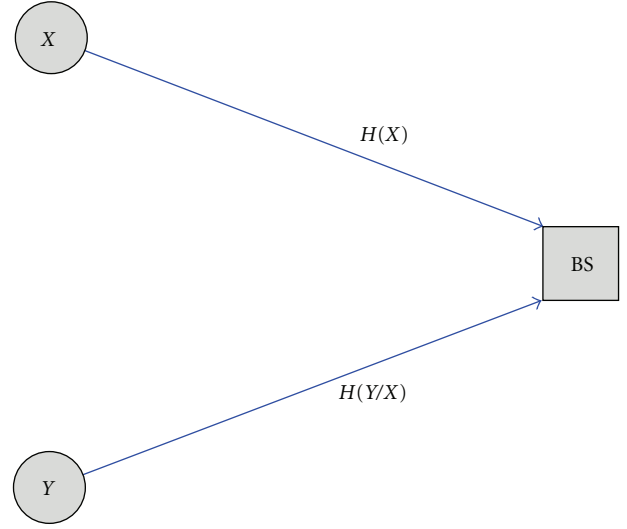


FIGURE 3: Two nodes WSN.

The node nearest to the BS has to encode at the rate of its entropy, while the second nearest node has to encode at the rate of the conditional entropy of itself given that the first source  $X_1$  is available at the BS, and so on. Let us call the solution of (8) the Minimum Total Power (MTP) optimization. This optimization does however not maximize the network lifetime. To illustrate this, consider the example depicted in Figure 3, with two correlated sources placed at an equal distance from the BS. The MTP solution will assign encoding rates according to (2). While this rate assignment minimizes the network's total power consumption, the network lifetime is not maximized since the nodes are encoding at different rates. In [24, 25], the authors searched to optimize the lifetime of the WSN using distributed source coding on cluster level with one-hop communication between cluster nodes and the cluster head. They consider the decoder to run at the cluster head.

### 3. Lifetime Optimization of WSN with Correlated Sensors

The above-mentioned code design, where side information is assumed available at the decoder, is called *asymmetric* Slepian-Wolf coding. For the network in Figure 3, the maximum network lifetime can be achieved by encoding at the *symmetric* Slepian-Wolf coding point, which is the green point shown in Figure 2. Most DSC designs can reach almost the theoretical limit by implementing *asymmetric* Slepian-Wolf coding [17–19, 26]. When using *asymmetric* Slepian-Wolf coding, the lifetime of the network shown in Figure 3 can be maximized by periodically switching coding rates between nodes every time interval  $T$ . If  $T$  is fixed, the lifetime of the network can be considered as the maximum multiple of this interval,  $mT$ ,  $m = \{1, 2, \dots, M\}$ , until the network is unable to perform its functionality successfully. Our goal is to maximize the lifetime of the network, which we can do by maximizing the value of  $M$ .

We model the WSN as a directed graph  $G = (X, E)$ , where the set  $X = \{X_1, X_2, \dots, X_N\}$  is the wireless nodes and  $t$  is the BS that collects and decodes the network data.  $E$  is the set of links connecting the nodes of set  $X$ . The edge  $(X_i, X_j)$  is an element of  $E$  if  $X_j$  is in the transmission range of  $X_i$ . The total number of links, that is, the size of  $E$ , is  $K$ . We further denote the initial energy and the current energy of node  $X_i$  by  $\text{IE}(X_i)$  and  $\text{CE}(X_i)$ , respectively. We allow the channel quality between nodes to change over time, so that for a link  $(X_i, X_j) \in E$ ,  $X_i$  requires  $e^m(X_i, X_j)$  energy to transmit one bit to node  $X_j$  at time index  $m$ . All nodes require a constant energy  $e_{rx}$  to receive one bit from any node. Node  $X_i$  is encoding its data using DSC at rate  $r_i^m$ . Since nodes also function as routers, we denote the rate of the data flow originally generated at node  $X_k$  forwarded from  $X_i$  to  $X_j$  in time slot  $m$  as  $r_{i,j}^{m,k}$ . Our goal is to maximize the number of time slots until the network is unable to deliver all nodes' data by optimizing the rate assignments:

$$\max_{r_i^m, r_{i,j}^{m,k}} M, \quad (9)$$

under the constraint of Slepian-Wolf encoding

$$\sum_{X_i \in X_s} r_i^m \geq H\left(\frac{X_s}{X_s^s}\right), \quad \forall X_s \subseteq X, \quad m = 1, 2, \dots, M. \quad (10)$$

The total flow at each node can be formulated as the difference between the output and the input data flows. If a node is a routing node, then the difference between its output and its input flows is zero. If the node is a source node, the difference between output and input flows is the rate of the node's encoded data:

$$\sum_{j \in X \setminus \{i\}} r_{i,j}^{m,k} - \sum_{j \in X \setminus \{i\}} r_{j,i}^{m,k} = \begin{cases} 0 & X_i \neq X_k, \\ r_i^k & X_i = X_k, \end{cases} \quad (11)$$

$$\forall X_i \in X, \quad X_k \in X, \quad \forall m = \{1, 2, \dots, M\}.$$

The relation between the data rate  $r_i^k$  and the transmission energy  $e^m(X_i, X_j)$  has to be found in order to formulate the energy consumption as a function to be optimized. Starting from Shannon's point-to-point wireless communication theorem [27], which states that the maximum transmission rate at which a transmitter can communicate its data to a receiver through an AWGN channel is bounded by the capacity of that channel, we have

$$R \leq B \log(1 + \text{SNR}), \quad (12)$$

where  $B$  is the channel bandwidth and SNR is the signal power to the noise power ratio at the receiver antenna. Without loss of generality, we ignored any interference effects from concurrent communications as well as other channel characteristics like fading and shadowing. (12) can be reformulated into

$$e^{R/B} - 1 \leq \text{SNR}. \quad (13)$$

Since WSN nodes have low data rates, the left-hand part of (13) can be approximated into

$$e^{R/B} (1 - e^{-R/B}) \approx e^{R/B} \left(\frac{R}{B}\right) \approx \frac{R}{B}. \quad (14)$$

From (14) and (13), we derive that the data rate  $r_i^k$  and the transmission power of  $X_i$  have a linear relation if the bandwidth  $B$  and the noise power at the receiving node are constant. From this linear relation, we deduce that the power consumed by a sensor node for transmission on a particular link is the unit power consumption for transmission on that link multiplied by the data rate of the link. Likewise, the power consumption for reception on a link is the product of the unit reception power consumption and the rate of incoming data on that link. Thus, the total energy consumption of the radio units of all nodes in the network is expressed by the following relation:

$$\sum_{m=1}^M \sum_{X_j \in X \setminus \{X_i\}} \sum_{X_k \in X \setminus \{X_i\}} (e_{Tx}^m(X_i, X_j) r_{i,j}^{m,k} + e_{Rx} r_{j,i}^{m,k}) \leq \text{IE}(X_i), \quad \forall X_i \in X. \quad (15)$$

From (9), (10), (11), and (15), we can see that lifetime optimization using DSC comprises two optimization problems: the rate optimization problem and the route optimization problem. This optimization problem is NP-hard since the routing optimization problem itself is NP-hard [28]. It is generally difficult to construct a routing tree that maximizes the network lifetime due to the involvement of two optimization objectives: maximizing the residual energy of each node and minimizing the network's total energy consumption. These two objectives are not necessarily complementary and might even conflict: a routing tree could, for example, minimize the network's total energy consumption by placing a high burden on a particular node. However, a routing algorithm, which uses link weights based on an exponential function of the network's resource utilization, has been shown to cope very efficiently with this optimization problem [29]. The authors of [29] assign to each edge a cost that is exponential in the currently occupied link capacity in order to optimize the throughput of the network. Furthermore, they derived bounds on the competitive ratio of their routing algorithm and proved that no other online routing algorithm can achieve a better competitive ratio. In [30], this routing algorithm was adapted to optimize the lifetime of Wireless Sensor Networks by updating the links' weights with the energy utilization of the nodes.

The authors of [31] use the same optimization criteria as in [30], and links are assigned cost functions which are exponential in the transmitter's energy utilization. In [32], algorithms based on the same exponential penalization are proposed to optimize the routing tree of heterogeneous networks, in which nodes differ in energy capacity. In all aforementioned related work, the energy consumed for the reception of data is neglected.

We developed two algorithms for optimizing the routes and the rates used in a WSN. The routing algorithm, which is an improvement of the CMAX algorithm described in [30], penalizes the network links according to an exponential function of the energy consumption for transmission and reception. Regarding the Slepian-Wolf rates optimization problem, it is known that for an optimal *symmetric* rate assignment in a network with more than 3 nodes the complexity of the decoder is difficult to implement practically [21]. The realization of the decoder becomes feasible if we allow the nodes to encode at *asymmetric* rates, that is, one node is decoded separately and its output is used as side information to decode the second node, then these two outputs are used as side information to decode the third node and so on. We developed the Rate Optimization (RO) algorithm which assigns asymmetric rates to the nodes. The algorithm first assigns the rates using MTP to minimize the network's total energy consumption, after which it performs a tradeoff between minimizing network energy and maximizing nodes' residual energy by swapping the rates of the nodes. The RO algorithm requires global knowledge of all nodes' rates and residual energies. Thus, the RO algorithm is centralized and is running at the BS, which broadcasts the updated rate assignment every periodic interval  $T$ . The routing algorithm is also executed at the BS every  $T$  seconds, and the new routes are broadcast together with the rates assignment.

**3.1. Routing and Rates Optimization Algorithms.** In order to explain our route optimization algorithm, we first describe how CMAX [30] works, on which our extension UCMA is built. After the BS collected the nodes' residual energies, the CMAX algorithm runs the following three steps.

*Step 1.* If all nodes have full energy (i.e.,  $CE(X_i) = IE(X_i)$ ), jump to Step 2 without modifying the graph  $G$ . Else, eliminate from  $G$  every edge  $e(X_i, X_j)$  for which  $CE(X_i) < e^m(X_i, X_j)$ , then change the weight of every remaining edge  $e^m(X_i, X_j)$  to  $e^m(X_i, X_j) \times (\lambda^{\alpha(X_i)} - 1)$ , where  $\alpha(X_i)$  is the energy utilization ratio of node  $X_i$ :

$$\alpha(X_i) = \frac{IE(X_i) - CE(X_i)}{IE(X_i)} = 1 - \frac{CE(X_i)}{IE(X_i)}, \quad (16)$$

where  $\lambda$  is a constant that quantifies the penalty of using a link.

*Step 2.* Find the shortest path between each node and the BS using Dijkstra's algorithm in the modified graph.

*Step 3.* Let  $\beta$  be the length of the shortest path found in Step 2 ( $\beta = \infty$  if no path was found). If  $\beta \leq \sigma$ , route the data along the shortest path, otherwise reject it.

The computational complexity of the CMAX algorithm is dominated by the shortest path computation (Step 2) and is  $O(K + N \log N)$ . The authors of [30] derived the competitive ratio of CMAX by comparing it to an optimal off-line routing algorithm. The competitive ratio of CMAX is found to be  $O(\log N \rho)$ , where  $\rho$  is the ratio of the edge with maximum

transmission energy to the edge with minimum transmission energy

$$\rho = \frac{\max_{i,j \in X} e(X_i, X_j)}{\min_{i,j \in X} e(X_i, X_j)}. \quad (17)$$

To find the competitive ratio,  $\lambda$  and  $\sigma$  are set to  $\lambda = 2(N\rho + 1)$  and  $\sigma = N \cdot \max_{i,j \in X} e(X_i, X_j)$ , respectively. Setting  $\sigma < \infty$  implies that packets may be rejected even if there is sufficient energy available to route the packet. Since our objective is to maximize the total number of packets delivered to the Base Station, we omit Step 3 in our modified routing algorithm, so that the route is not to be rejected if there is enough energy to deliver a packet over it.

Most WSN nodes, that are available on the market today, consume more energy while in receive mode rather than in transmit mode, even when the node is transmitting at the maximum power. The widely used transceiver chip CC2420 [33] consumes 18.8 mA in the receive mode, while it consumes 17.4 mA in the transmit mode at maximum transmission power. The authors of [30–32] do not take into account the energy spent in the receive mode while optimizing the routing tree. We updated CMAX to include the reception costs by modifying the weights of the graph's edges. The Updated-CMAX (UCMAX) runs the following steps.

*Step 1.* If all nodes have full energy (i.e.,  $CE(X_i) = IE(X_i)$ ), jump to Step 2 without modifying the graph  $G$ . Else, eliminate from  $G$  every edge  $e(X_i, X_j)$  for which  $CE(X_i) < e^m(X_i, X_j)$ , then change the weight of every remaining edge to

$$e(X_i, X_j) \rightarrow e(X_i, X_j) \times (\lambda^{\alpha(X_i)} - 1) + e_{Rx} \times (\gamma^{\alpha(X_j)} - 1), \quad (18)$$

where  $\lambda$  and  $\gamma$  are constant parameters that quantify the penalty of using the link  $e(X_i, X_j)$  based on the energy utilization of the transmitting node  $X_i$  and the receiving node  $X_j$ .

*Step 2.* Find the shortest path between each sensor node and the BS using Dijkstra's algorithm in the modified graph.

UCMAX avoids to route network data through nodes with low residual energy. The Rate Optimization algorithm (RO) runs on top of the optimized routing tree found by UCMA as follows.

*Step 1.* Assign the rates to the nodes according to MTP using the routing tree found by UCMA.

*Step 2.* Calculate the total energy consumption of the network

$$P_L = \sum_{i=1}^N w_i^m r_i^m, \quad (19)$$

where  $w_i^m$  is the total energy required to route one bit from node  $X_i$  to the BS, and  $r_i^m$  is the rate of node  $X_i$  during time slot  $m$ .

*Step 3.* Find the node with the minimum residual energy, let it be  $X(\min)$ .

*Step 4.* Search for another node with lower data rate and higher residual energy and name it  $X(tmp)$ . The two nodes should not be on the same routing path to the BS.

*Step 5.* Swap the rates between  $X(\min)$  and  $X(tmp)$ , so that if  $X(\min)$  is encoding at rate  $H(X_1)$  and  $X(tmp)$  at rate  $H(X_2/X_1)$ ,  $X(\min)$ 's rate should become  $H(X_1/X_2)$  while  $X(tmp)$ 's rate should become  $H(X_2)$ .

*Step 6.* Calculate the total network power with (19) and store it in a vector  $P$ . Go back to Step 4 and repeat for all possible rate switches. Choose the rate swap with the minimum total power in  $P$  and let us name the new total power  $P_L^{new}$ .

*Step 7.* If  $P_L^{new}$  is less than  $zP_L$ , where  $z$  is a constant parameter, accept the new rates. Else, do not switch the rates.

*Step 8.* Go back to Step 3 and repeat until all possible rate switches are tested for all nodes.

Parameter  $z$  allows our optimization to balance between minimum total network energy and maximum per node energy. At  $z = 1$ , the network retains the minimum total network energy achieved by MTP while trying to maximize per node energy. For  $z < 1$ , the RO algorithm is simply the MTP algorithm.

## 4. Simulations and Results

*4.1. Experimental Setup.* MatLab simulations are used to evaluate the performance of the optimization algorithm. We simulate an environment area of  $100 \times 100 \text{ m}^2$  and consider two network configurations: in the first one nine nodes are located on a  $3 \times 3$  grid, while in the second one the set of nodes is extended to twenty-five nodes arranged in a  $5 \times 5$  grid. The BS is located at the center in both grids. The two grids allow us to compare the performance of the optimization algorithms with different node densities. The energy consumption of the nodes is approximated by the energy consumed by the radio transceiver. The energy for reception,  $e_{Rx}$ , is considered the same for all nodes. Without loss of generality, the energy for transmission  $e_{Tx}^m(X_i, X_j)$  is assumed to depend on the distance between the nodes  $X_i$  and  $X_j$ . We use the following model:

$$e_{Tx}(X_i, X_j) = \left[ \beta d(X_i, X_j)^\kappa + \rho \right] \times T_{Tx}, \quad (20)$$

where  $\beta$ ,  $\kappa$ , and  $\rho$  are constants and their values depend on the radio chip's characteristics and the environmental conditions.  $T_{Tx}$  is the packet transmission time. The channel parameters used in our simulations are shown in Table 1 and are calculated according to the work of [34]. The correlation between the sensed data at the nodes is represented by a Gaussian model [22]

$$f(x) = \frac{1}{\sqrt{2\pi} \det(K)^{1/2}} e^{-((1/2)(X-\mu)K^{-1}(X-\mu))}, \quad (21)$$

TABLE 1: Experiment parameters.

$\beta = 5.219 \times 10^{-4}$	$\sigma = 1$
$\kappa = 3.5$	$c = 0.001$
$\rho = 1.2 \times 10^{-5}$	Bit rate = 250 kb/s
Transmission range = 100 m	Maximum packet size = 128 bytes
$IE(X_i) = 10 \text{ Joule}$	Maximum $T_{Tx} = 4.1 \text{ ms}$
$e_{Rx} = 59.1 \times 10^{-3} \times T_{Tx} \text{ Joule}$	

where  $K$  is the covariance matrix which represents the spatial correlation between the measurements.  $K$  is created by assuming that these correlations are changing according to the distance between the nodes. More precisely, the following model  $K_{i,j} = \sigma^2 \exp(-c|d(X_i, X_j)|^2)$  is used to define this relationship, where  $K_{i,j}$  and  $d(X_i, X_j)$  represent the correlation and the distance between the nodes  $X_i$  and  $X_j$ , respectively.  $\sigma^2$  is the variance of the nodes' sensed data (we consider all nodes to have the same variance) and  $c$  is the attenuation factor of the correlation between the nodes.

*4.2. Simulation Results.* The role of  $\lambda$  in the CMAX algorithm is studied in [30], where it acts as a penalty factor for using the links between nodes with low residual energy in the WSN. For  $\lambda = 1$ , the routing structure is the same as the Shortest Path Tree. Increasing  $\lambda$  improves the per node residual energy at the expense of a higher total network energy consumption. It is shown in [30] through experiments that at  $\sigma = \infty$  and for large values of  $\lambda$ , the number of total delivered messages is maximized and becomes insensitive to increasing values of  $\lambda$ . In our simulation, we set  $\sigma = \infty$  and  $\lambda = 10,000$  for all experiments to analyze the effect of other factors on the lifetime of the network.

*4.2.1. UCMAx versus CMAx.* For both the  $3 \times 3$  and  $5 \times 5$  grids, we execute several runs of the routing algorithm while changing the value of  $\gamma$  at each run. The rates of the nodes are assigned with MTP. The update period (the total number of measurement collection cycles to the Base Station before the algorithm calculates a new routing tree)  $T$  is set to 1000. Figure 5 presents the network lifetime (total number of route updates until there is no possible route to deliver all network data) on the  $y$ -axis and  $\gamma$  on the  $x$ -axis. The network lifetime of CMAX is constant since  $\lambda$  is constant and the CMAX algorithm is not affected by changes in  $\gamma$ . In the simulations, we include the reception energy utilization in the calculations of the edges' new weights. As shown in Figure 5, the  $3 \times 3$  network has a longer lifetime than the  $5 \times 5$  network with the CMAX routing optimization, even though the  $5 \times 5$  network is denser. The  $5 \times 5$  network has a shorter lifetime because the CMAX algorithm does not take into account energy consumed during receive mode, which can be much higher than the energy used in transmit mode in a dense network. As we pointed out before, for most WSN nodes' transceivers, the power consumption in receive mode is higher than the transmission energy utilization, even at the maximum transmission power. In the  $3 \times 3$  network, the nodes consume more energy in the transmit mode compared

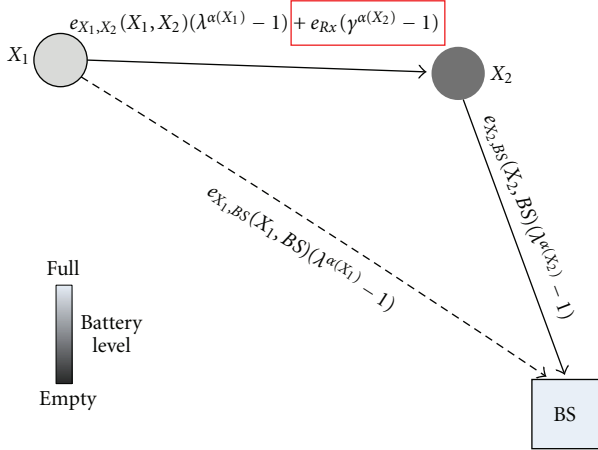


FIGURE 4: A two-node network with a Base Station.

to the nodes in the  $5 \times 5$  network which results in a better optimization of the routing tree for the  $3 \times 3$  network compared to the  $5 \times 5$  network when using CMAX. The large gap between the estimated power consumption of CMAX and the calculated power consumption using (20) leads to a shorter lifetime for the  $5 \times 5$  network.

The lifetime of the network varies with  $\gamma$  in the UCMAX algorithm. At  $\gamma = 1$ , UCMAX performs the same as CMAX. However by increasing  $\gamma$ , the lifetime of the network improves and reaches a maximum at  $\gamma = 10$  for both the  $3 \times 3$  and the  $5 \times 5$  networks. The optimum value of  $\gamma$  depends on the parameters used in the energy consumption model of (20).

To describe the advantage of UCMAX over CMAX, let us consider the network in Figure 4. Using the CMAX optimization algorithm, node  $X_1$  may choose to route its data through  $X_2$  if the sum of the weights of the links  $(X_1, X_2)$  and  $(X_2, BS)$  is less than the weight of link  $(X_1, BS)$ . Recall that CMAX neglects the energy consumption for reception at  $X_2$ . By including this reception energy utilization of  $X_2$  in the weight of the edge  $(X_1, X_2)$ , UCMAX can decide to avoid routing through  $X_2$ , since  $X_2$  pays a double price in terms of energy, that is, for receiving and transmitting.

With UCMAX, the  $5 \times 5$  network, which is denser than the  $3 \times 3$  network, has a longer lifetime, since in dense networks the distance between nodes is shorter and thus the nodes require less transmission power.

**4.2.2. RO Algorithm Evaluation.** To compare the RO algorithm to MTP, we applied the RO algorithm on top of UCMAX.  $\lambda$  and  $\gamma$  are set to  $\lambda = 10,000$  and  $\gamma = 10$ , respectively. Figure 6 depicts the variations in the network lifetime while changing the parameter  $z$ . For  $z < 1$ , there is no improvement in the lifetime when compared to MTP. For  $z \geq 1$ , the RO algorithm shows large improvements in network lifetime for both the  $5 \times 5$  and  $3 \times 3$  networks. The improvement in the lifetime at  $z = 1$  is due to the the grid structure of the networks with the BS positioned at the center. In a grid network, some nodes are at an equal distance from the BS. When assigning different rates to these

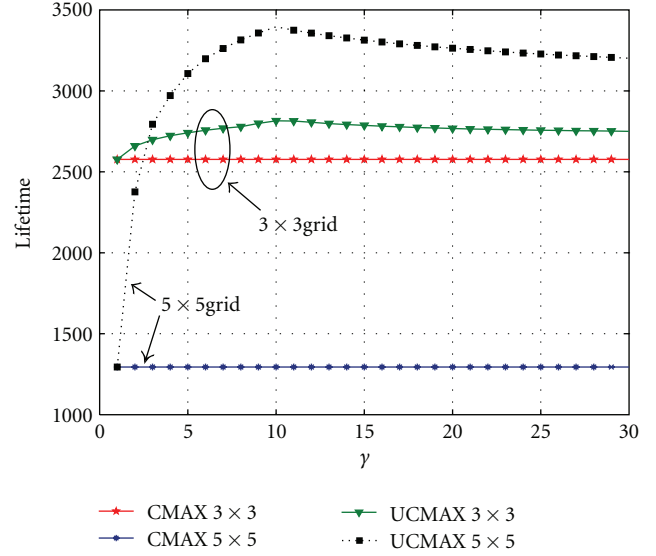


FIGURE 5: Routing optimization for maximizing the network lifetime.

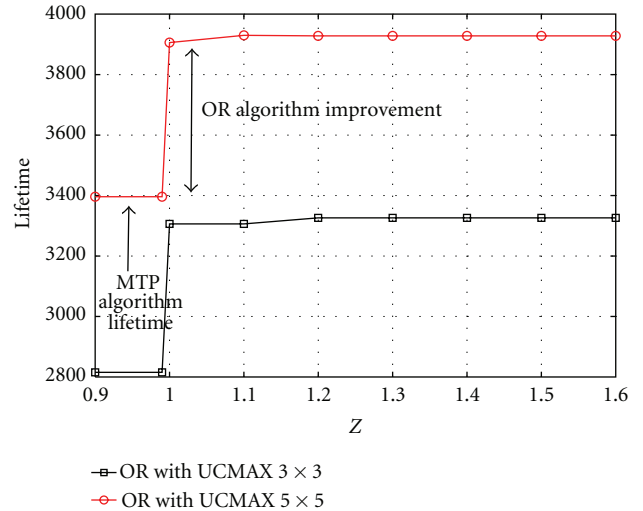


FIGURE 6: Network lifetime with optimum rate algorithm.

nodes, the total network power  $P_L$  will remain the same. After each update period  $T$ , the RO algorithm switches the rates between the nodes with equal distance to the BS. Low rates are assigned to nodes with minimum residual energy while healthier nodes are penalized with higher rates. The network lifetime is insensitive to high values of  $z$  because assigning low rates to nodes with low residual energy close to the BS and high rates to more distant nodes is not improving the lifetime of the network, since the nodes close to the BS always need to route the data from the farther nodes.

**4.2.3. Update Period  $T$ .** In Figure 7 we show the effect of the update frequency of the routing tree and the encoding rates on the lifetime of the network. Before each update, the

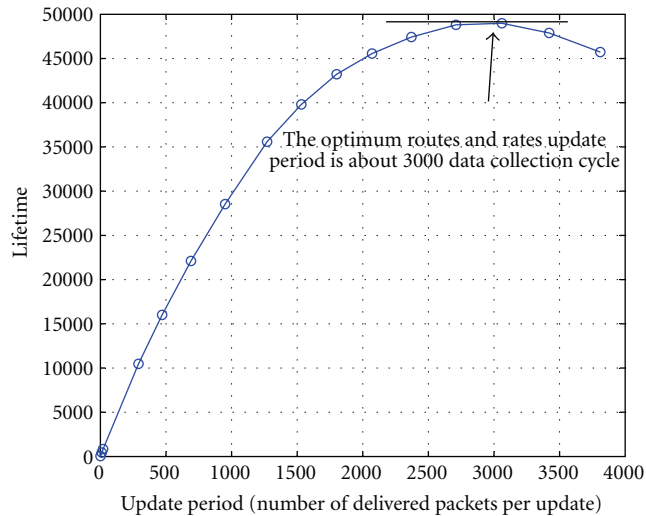


FIGURE 7: Effect of update period  $T$  on the network lifetime.

nodes need to communicate their residual energy to the BS, and the BS subsequently broadcasts the optimized routing tree and rates back to the nodes. We assume that the BS can encapsulate the routing tree and the rates of the nodes in one broadcast packet. When the BS broadcasts the packet, the nearest nodes receive the packet and then forward it to the nodes further in the network and so on. We consider that each node consumes energy equal to the reception and the transmission of one packet during the broadcast process and we neglect the energy spent in collecting the nodes' residual energies. By running the UCMA and RO algorithms on the  $5 \times 5$  network with different update periods  $T$ , we found that at  $T = 3000$  the network has the longest lifetime. At fast updating rates, the communication overhead caused by the broadcasts reduces the network lifetime, while at slow updating rates, the algorithm does not track accurately the depletion rate of the batteries of the nodes.

## 5. Conclusion

In this paper we considered Wireless Sensor Networks placed in a specific geographical area, gathering correlated information from multiple nodes that forward their data to a Base Station. We addressed the maximization of network lifetime through the application of distributed source coding for the compression of the spatially correlated data. The motivation for using DSC instead of Explicit Communication is the possibility of decomposing the optimization problem in two separate problems: optimizing the routes and the rates independently. The paper presents two algorithms: the first one is a routing optimization algorithm and the second one is a rate optimization algorithm. The first algorithm, the Updated-CMAX algorithm (UCMA), improves the CMAX algorithm, as presented in the literature, by taking into account the energy utilization of nodes in receive mode. The second algorithm, the Rate Optimization algorithm (RO), balances between minimizing total network energy consumption and the per node energy consumption while

assigning Slepian-Wolf encoding rates. The RO algorithm assigns low rates to nodes with low residual energy and higher rates to nodes with excessive energy.

Our experiments show that UCMA provides a significant improvement in terms of network lifetime in comparison to CMAX. With respect to the minimum total network Slepian-Wolf rates, our RO algorithm improved lifetime by 17%. The two algorithms combined provide a full-optimized solution that maximizes the lifetime of networks collecting correlated data. The optimal update period between successive rate and route optimizations is also derived by taking into account the broadcasting energy consumption needed to send the optimized rate and route assignments from the Base Station.

## Acknowledgments

The work presented in this paper was supported by EMECW (Erasmus Mundus External Corporation Window lot3) and the FWO (Fonds Wetenschappelijk Onderzoek), project G.0219.09N.

## References

- [1] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer Networks*, vol. 52, no. 12, pp. 2292–2330, 2008.
- [2] G. Anastasi, M. Conti, M. di Francesco, and A. Passarella, "Energy conservation in wireless sensor networks: a survey," *Ad Hoc Networks*, vol. 7, no. 3, pp. 537–568, 2009.
- [3] G. Barrenetxea, F. Ingelrest, G. Schaefer, M. Vetterli, O. Couach, and M. Parlange, "SensorScope: out-of-the-box environmental monitoring," in *Proceedings of the International Conference on Information Processing in Sensor Networks (IPSN '08)*, pp. 332–343, April 2008.
- [4] C. Hua and T. S. P. Yum, "Optimal routing and data aggregation for maximizing lifetime of wireless sensor networks," *IEEE/ACM Transactions on Networking*, vol. 16, no. 4, pp. 892–903, 2008.
- [5] Q. Gao, K. J. Blow, D. J. Holding, and I. Marshall, "Analysis of energy conservation in sensor networks," *Wireless Networks*, vol. 11, no. 6, pp. 787–794, 2005.
- [6] G. Xing, C. Lu, Y. Zhang, Q. Huang, and R. Pless, "Minimum power configuration for wireless communication in sensor networks," *ACM Transactions on Sensor Networks*, vol. 3, no. 2, Article ID 1240231, 2007.
- [7] K. Sha and W. Shi, "Modeling the lifetime of wireless sensor networks," *Sensor Letters*, vol. 3, no. 2, pp. 126–135, 2005.
- [8] S. Patten, B. Krishnamachari, and R. Govindan, "The impact of spatial correlation on routing with compression in wireless sensor networks," *ACM Transactions on Sensor Networks*, vol. 4, no. 4, pp. 1–33, 2008.
- [9] R. Cristescu, B. Beferull-Lozano, M. Vetterli, and R. Wattenhofer, "Network correlated data gathering with explicit communication: NP-completeness and algorithms," *IEEE/ACM Transactions on Networking*, vol. 14, no. 1, pp. 41–54, 2006.
- [10] H. Luo and G. J. Pottie, "Designing routes for source coding with explicit side information in sensor networks," *IEEE/ACM Transactions on Networking*, vol. 15, no. 6, pp. 1401–1413, 2007.



- [11] H. Wang, D. Peng, W. Wang, and H. Sharif, "Optimal rate-oriented routing for distributed source coding in wireless sensor network," in *Proceedings of the 2nd ACM International Workshop on Quality of Service and Security in Wireless and Mobile Networks (Q2SWinet '06)*, pp. 55–58, ACM Press, New York, NY, USA, 2006.
- [12] D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive codes for distributed source coding," *Signal Processing*, vol. 86, no. 11, pp. 3123–3130, 2006.
- [13] A. Kashyap, L. A. Lastras-Montañó, C. Xia, and Z. Liu, "Distributed source coding in dense sensor networks," in *Proceedings of the Data Compression Conference*, no. 2, pp. 13–22, IEEE, 2005.
- [14] D. Marco and D. L. Neuhoff, "Reliability vs. efficiency in distributed source coding for field-gathering sensor networks," in *Proceedings of the 3rd International Symposium on Information Processing in Sensor Networks (IPSN '04)*, pp. 161–168, New York, NY, USA, ACM Press, 2004.
- [15] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973.
- [16] S. S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 51–60, 2002.
- [17] J. Bajcsy and P. Mitran, "Coding for the Slepian-Wolf problem with turbo codes," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '01)*, vol. 2, pp. 1400–1404, IEEE, 2001.
- [18] A. Aaron and B. Girod, "Compression with side information using turbo codes," in *Proceedings of the Data Compression Conference*, 2002.
- [19] A. D. Liveris, Z. Xiong, and C. N. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Communications Letters*, vol. 6, no. 10, pp. 440–442, 2002.
- [20] C. F. Lan, A. D. Liveris, K. Narayanan, Z. Xiong, and C. Georghiades, "Slepian-Wolf coding of multiple M-ary sources using LDPC codes," in *Proceedings of the Data Compression Conference (DCC '04)*, p. 549, March 2004.
- [21] A. D. Liveris, C. F. Lan, F. Narayanan, Z. Xiong, and C. N. Georghiades, "Slepian-Wolf coding of three binary sources using LDPC codes," in *Proceedings of the International Symposium on Turbo Codes and Related Topics*, pp. 1–4, 2003.
- [22] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "Networked Slepian-Wolf: theory, algorithms, and scaling laws," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4057–4073, 2005.
- [23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, vol. 306 of *Wiley Series in Telecommunications*, Wiley-Interscience, New York, NY, USA, 1991.
- [24] H. Arjmandi, M. Taki, and F. Lahouti, "Lifetime maximized data gathering in wireless sensor networks using limited-order distributed source coding," *Signal Processing*, vol. 91, no. 11, pp. 2661–2666, 2011.
- [25] H. Arjmandi and F. Lahouti, "Resource optimized distributed source coding for complexity constrained data gathering wireless sensor networks," *IEEE Sensors Journal*, vol. 11, no. 9, Article ID 5705538, pp. 2094–2101, 2011.
- [26] D.S. Lun, M. Médard, T. Ho, and R. Koetter, "Network coding with a cost criterion," in *Proceedings of the International Symposium on Information Theory and its Applications (ISITA '04)*, pp. 1232–1237, Citeseer, April 2004.
- [27] C. E. Shannon, "The mathematical theory of communication," *M.D. Computing : Computers in Medical Practice*, vol. 14, no. 4, pp. 306–317, 1948.
- [28] Q. Li, J. Aslam, and D. Rus, "Online power-aware routing in wireless Ad-hoc networks," in *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking (MOBICOM '01)*, pp. 97–107, 2001.
- [29] B. Awerbuch, Y. Azar, and S. Plotkin, "Throughput-competitive on-line routing," in *Proceedings of the 34th Annual Symposium on Foundations of Computer Science*, pp. 32–40, IEEE Computer Society, Washington, DC, USA, 1993.
- [30] K. Kar, M. Kodialam, T. V. Lakshman, and L. Tassiulas, "Routing for network capacity maximization in energy-constrained ad-hoc networks," in *Proceedings of the 22nd Annual Joint Conference of the IEEE Computer and Communications (IEEE INFOCOM '03)*, vol. 1, pp. 673–681, IEEE Societies, 2003.
- [31] W. Liang and Y. Liu, "Online data gathering for maximizing network lifetime in sensor networks," *IEEE Transactions on Mobile Computing*, vol. 6, no. 1, pp. 2–11, 2007.
- [32] L. Lin, N. B. Shroff, and R. Srikant, "Asymptotically optimal energy-aware routing for multihop wireless networks with renewable energy sources," *IEEE/ACM Transactions on Networking*, vol. 15, no. 5, pp. 1021–1034, 2007.
- [33] "CC2420 2.4 GHz IEEE 802.15.4 / ZigBee-ready RF transceiver," Tech. Rep., Texas Instruments, 2007.
- [34] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Transactions on Wireless Communications*, vol. 4, no. 5, pp. 2349–2360, 2005.