

# EFFICIENT DISTRIBUTED MULTIREOLUTION PROCESSING FOR DATA GATHERING IN SENSOR NETWORKS

*Jugoslava Acimović, Răzvan Cristescu and Baltasar Beferull-Lozano*

Audiovisual Communications Laboratory  
Swiss Federal Institute of Technology in Lausanne – EPFL  
Lausanne CH-1015, Switzerland

{Jugoslava.Acimovic, Razvan.Cristescu, Baltasar.Beferull}@epfl.ch

## ABSTRACT

We consider large sensor networks where the cost of collecting data from the network nodes to the data gathering sink is critical. We propose several algorithms that use limited local communication and distributed signal processing to make communication more efficient in terms of transmission cost. We consider a model that uses distributed wavelets-based signal processing. We first propose an algorithm that performs processing at nodes as data is forwarded to the sink. Then, we analyze algorithms that perform network division into groups of adaptive size and for which signal processing is applied separately to each group. We show by numerical simulations that such multiresolution approaches result in significant improvements for data gathering in terms of total communication costs.

## 1. INTRODUCTION

Consider a network of sensors that measure certain data and have to transmit all obtained information to a common central node, or sink, for processing or storage. Each sensor is equipped with a limited autonomous battery that provides energy for the signal transmission. Usually, in sensor networks, the cost of communication is substantially higher than the cost of local processing [5]. Since the sensors are strongly constrained in terms of battery power [5], a meaningful task in such settings is to design algorithms for data transmission that minimize the total cost of gathering the measured data [1, 4, 6]. The simplest possible strategy consists of direct transmission of the information from all sensors to the sink, and we refer to this as raw data gathering. However, it is clear that even for moderately sized networks this approach results in very inefficient use of resources.

Several features can be exploited to decrease the total cost of communication. The measured phenomenon is typically a spatially dependent continuous process, and the data obtained by the sensors represent its space-sampled version. Consequently, physically close nodes measure correlated data, and this similarity can be exploited to decrease the amount of information needed by the sink for the data reconstruction. Distributed signal processing in intermediate nodes may take advantage of this property by performing signal decorrelation between the sensors, thus reducing the total cost of transmission. Further improvements can be obtained by using multi-hop communication, having sensors far from the sink relay their data via neighboring nodes rather than sending directly to the sink.

The related problem where the measured data are random variables and conditional entropy coding is employed is studied in

[1]. In this work, we focus on a specific class of deterministic signals for which wavelet processing is appropriate, namely piecewise defined signals, such as for instance piecewise constant signals, which can be often found in practice after some threshold detection process. For deterministic piecewise continuous signals, the amount of correlation present in the signal is essentially determined by the number of discontinuities. We will analyze how the number of discontinuities affects the total transmission cost when signal decorrelating is employed for data gathering, in comparison to raw data gathering. A joint consideration of wavelet coding and energy efficient transmission is exploited in [2], where wavelet lifting is used. However, the signal model they consider is different.

We analyze several wavelet-based algorithms for the task of efficient data gathering. First, we consider a distributed algorithm applied on the whole network globally: the measured process is approximated successively with different resolutions as data is sent to the sink, by calculating low-pass (LP) and high-pass (HP) Haar transform coefficients in certain network nodes. Next, we propose an algorithm that involves network segmentation into groups of equal size. Distributed processing and local data gathering are performed inside each group separately, then the full set of coefficients from each group is transmitted to the sink. Finally, we propose an algorithm in which the network division into groups adapts to the characteristics of the specific measured signal. We show by numerical simulations that our algorithms significantly improve data gathering in terms of cost over raw data gathering. Also, the adaptive algorithm provides a further improvement of 20% over the fixed size group segmentation, for large networks. Moreover, we show how the improvements in terms of cost increase with the decrease of the number of discontinuities present in the signal.

In Section 2, we describe our sensor network scenario and assumed signal models. In Section 3 we introduce an algorithm for multiresolution data gathering using the Haar transform. In Section 4, we study the improvements of this algorithm when *local* group data gathering is employed. Finally, in Section 5 we present simulation results that compare the described algorithms. We conclude with Section 6.

## 2. PROBLEM STATEMENT

### 2.1. Network and Signal Model

For the sake of simplicity, we consider a one-dimensional<sup>1</sup> network model. The sensors, denoted as  $1, 2, \dots, N$ , and the network sink  $N + 1$  are placed equidistantly on a line (see Fig. 1). Without

<sup>1</sup>We are currently studying the extension to two-dimensional networks.



**Fig. 1.** A one-dimensional network model. The sink, denoted  $N + 1$ , is placed at the right extremity.

loss of generality, assume that the network size is  $N = 2^M$ , where  $M$  is a positive integer.

We consider a particular class of signals, namely continuous piecewise constant signals with a finite number of (uniformly distributed) discontinuities. This type of signals appears in many practical problems, for instance in thresholding/alert scenarios. Between each two discontinuities, the signal takes a value in the interval  $[a, b]$ , with  $a, b$  real numbers. The whole set of measurements at the network nodes is a vector  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)$  where  $Y_i$  corresponds to the sample obtained by sensor  $i$ . Thus, the vector  $\mathbf{Y}$  represents the space-sampled version of the original signal with sampling step  $d$  along the line (note that both space sampling and quantization introduce distortion into the reconstructed signal [3]). In this work, we assume that there is no entropy coding by the sensors, that is, for a given quantization step size  $\Delta$ , the coefficient values obtained by the sensors are coded using  $c = \log \left\lceil \frac{|a-b|}{\Delta} \right\rceil$  bits. The same quantization step size  $\Delta$  is used for all transform coefficients.

## 2.2. Haar Transform and Signal Approximation

Without loss of generality, we consider the  $M$ -level Haar wavelet transform, known to provide an efficient data representation for piecewise constant processes (this can be easily generalized to the case of piecewise polynomial signals by using the appropriate higher-order wavelet processing). For each level  $k = 1 \dots M$  of the transform, LP and HP coefficients are computed in the usual manner as sum and differences of previous lower level coefficients.

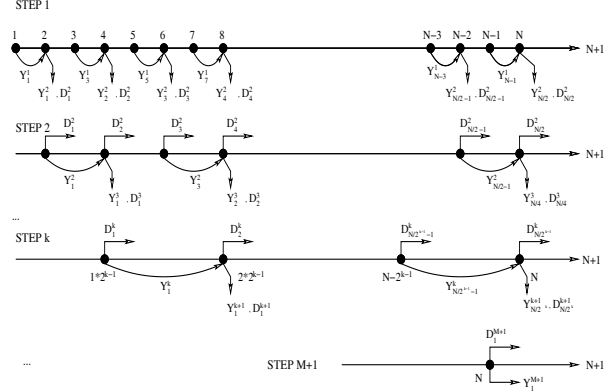
The set of transform coefficients is sufficient for signal reconstruction in the network sink (the signal reconstruction is only approximate). Notice that for the intervals where the signal is constant, all HP coefficients are zero. We assume that the power required to transmit zero-valued coefficients is negligible.<sup>2</sup> Therefore, the only information that needs to be transmitted are the LP coefficients. If the process has a single discontinuity then there is at most one non-zero HP coefficient in each multiresolution level  $k$ , which results in at most  $M$  non-zero HP coefficients for the whole representation. Generally, for a signal with  $P$  discontinuities, an upper bound for the number of non-zero HP coefficients (i.e. coefficients that have to be transmitted) is given by  $\min(P \cdot M, N)$ .

## 2.3. Cost Function and Reference Algorithm

We assume a linear dependence of the cost on the square of the distance [5]. Denote the bit-rate for coding data  $\xi$  as  $R(\xi)$  and the distance between transmitter  $X$  and receiver  $Y$  as  $d(X, Y)$ . Then, the cost function for sending the data between  $X$  and  $Y$ :

$$\text{cost}(X \rightarrow^\xi Y) = R(\xi) \cdot d^2(X, Y) \quad (1)$$

<sup>2</sup>This assumption is approximately valid for a coding scheme that uses a very small number of bits for coding zero-valued, compared to the number of bits used for coding the other coefficients. For instance, in [7] we use a simple prefix-coding approach with one bit sent for each zero HP coefficient.



**Fig. 2.** Haar transform multiresolution processing.

The total cost of a data gathering algorithm is the summation of (1) over all transmitter-receiver pairs in all the steps of the algorithm. Note that when multihop transmissions are considered, the total cost depends only on the number of hops, and not on the power of the distance.

We compare our algorithms with a simple raw data gathering scenario used as a reference model, in which each sensor sends the measured signal directly to the sink. In this case also, all signals are quantized before transmission and  $c$  bits are used for coding each of them. For this reference model, the total cost for a network with  $N$  sensors is  $\text{cost}(\text{ref}) = c \frac{N \cdot (N+1)}{2} d^2$ . For the evaluation of the performance of an algorithm  $A$ , we compute the relative cost:

$$\eta = \frac{\text{cost}(A)}{\text{cost}(\text{ref})}.$$

## 3. MULTIRESOLUTION DATA GATHERING USING A DISTRIBUTED HAAR TRANSFORM

The first algorithm that we propose, called *NetHaar*, uses a simple multiresolution processing based on the multiresolution Haar transform (see Fig. 2).

Denote  $Y_i^1 = Y_{2i-1}$ ,  $i = 1 \dots N/2$ . In the first step of the algorithm, sensors with odd indices send their data to the sensors with even indices. The transmitted signals are  $\mathbf{Y}^1 = (Y_1^1, Y_2^1, \dots, Y_{N/2}^1)$ . Next, sensors  $2, 4, \dots, N$  compute the Haar LP coefficients  $\mathbf{Y}^2 = (Y_1^2, Y_2^2, \dots, Y_{N/2}^2)$  and HP coefficients  $\mathbf{D}^2 = (D_1^2, D_2^2, \dots, D_{N/2}^2)$  by using the set of received measurements and their own measurements. Since sensors with even indices have all information necessary for an approximate reconstruction of the measured signal, we can ignore sensors with odd indices in further analysis. The next algorithm steps implement a similar procedure. Namely, in the  $k^{\text{th}}$  step,  $k = 1 \dots M$ , nodes  $2^{k-1} \cdot 1, 2^{k-1} \cdot 3, \dots, N - 2^{k-1}$  send their LP coefficients to nodes  $2^{k-1} \cdot 2, 2^{k-1} \cdot 4, \dots, N$ , and simultaneously their respective HP coefficients to the sink (see Fig. 2). Then, coefficients  $Y_i^{k+1}$  and  $D_i^{k+1}$  are computed as:

$$Y_i^{k+1} = \frac{Y_{2i-1}^k + Y_{2i}^k}{\sqrt{2}}, D_i^{k+1} = \frac{Y_{2i-1}^k - Y_{2i}^k}{\sqrt{2}}$$

for  $k = 1, 2, \dots, M$  and  $i = 1, 2, \dots, \frac{N}{2^k}$ .

Finally, in the  $M + 1$ -th step, node  $N$  sends both coefficients calculated in the previous step,  $Y_1^{M+1}$  and  $D_1^{M+1}$ , to the sink. At

the termination of this algorithm the sink has all data necessary for reconstructing an approximation of the signal.

Two types of transmissions occur, namely sensor-to-sensor and sensor-to-sink transmission. Thus, the total cost is:

$$\text{cost} = \text{cost}_s + \text{cost}_{sd}$$

where  $\text{cost}_s$  is the total sensor-to-sensor communication cost, and  $\text{cost}_{sd}$  is the total sensor-to-sink cost:

$$\text{cost}_s = \sum_{k=1}^M \sum_{i=1}^{\frac{N}{2^k}} R(Y_i^k) 2^{k-1} d^2,$$

$$\text{cost}_{sd} = \sum_{k=2}^{M+1} \sum_{i=1}^{\frac{N}{2^{k-1}}} R(D_i^k) (N - 2^{k-1}i + 1) d^2 + R(Y_1^{M+1}) d^2,$$

since every transmitter in the  $k^{\text{th}}$  step sends its LP coefficient over  $2^{k-1}$  hops and its HP coefficient over  $N - 2^{k-1}i + 1$  hops.

By substituting the bit rates for coding transform coefficients, given in sections 2.1 and 2.2, into the expression for the cost function, we obtain the scaling laws for several illustrative examples (for the proof, see [7]). Namely, for a signal with one or no discontinuities, we have  $\eta \sim O\left(\frac{(\log_2 N)^2}{N}\right)$ , therefore, the relative cost  $\eta \rightarrow 0$  when  $N \rightarrow \infty$ . On the other hand, for a signal with all non-zero HP coefficients we have  $\eta \rightarrow 1 + \frac{1}{e}$  for  $N \rightarrow \infty$ . Thus, for a sufficiently small number of non-zero HP coefficients (i.e. signal discontinuities) the *NetHaar* algorithm provides a gain, but when the number of such coefficients increases, its performance is limited. However, by limiting the number of discontinuities present in each set of signal samples, additional decreasing of the total cost can be obtained. This idea is further exploited in the algorithms described in the next section.

#### 4. DISTRIBUTED ALGORITHMS WITH NETWORK SEGMENTATION AND HAAR TRANSFORM

In this section, we introduce a set of algorithms based on multiresolution distributed processing which additionally adapts to the measured signal, while not necessarily requiring local processing at *all* nodes. The main idea behind these algorithms is to reduce sensor-to-sink communication by introducing an initial step of local communication between sensors, at the level of *groups* whose size will depend on the distribution of discontinuities. For instance, in the case of piecewise-constant signals, a large percentage of the HP coefficients are zero. Since the power for sending zero-valued coefficients is negligible, it is preferable to send these HP coefficients directly to the sink and to send LP coefficients between sensors, over shorter distances. A set of sensors is called a *group* if one of the sensors, the *leader-node*, collects all necessary local information to represent the signal data measured by all the sensors in the spatial region covered by the group.

For the sake of clarity in the presentation of the algorithms, consider the  $M + 1$  level binary tree representation of the distributed network (see Fig. 3). Each tree leaf corresponds to single-sensor groups, while the tree root contains the whole network. Each in-tree node represents a group of sensors.

The possible ways of communication between groups proposed in our algorithms are shown with arrows in Fig. 3. For instance, after the transmissions in the first step of the *NetHaar* algorithm, 2-sensor groups  $(2i - 1, 2i), i = 1, 2, \dots, \frac{N}{2}$  are created. After the

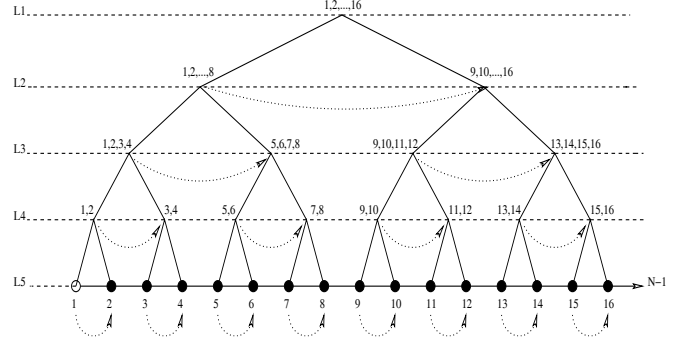


Fig. 3. Tree representation of the algorithms.

second step of the algorithm, 4-sensor groups are created:  $(4i - 3, 4i - 2, 4i - 1, 4i), i = 1, 2, \dots, \frac{N}{4}$ , and so on.

The *NetHaar* algorithm fully spans the tree from bottom to top, without checking for optimality of this operation in terms of cost. Moreover, it requires that all sensor nodes have signal processing abilities (to perform the Haar transform). The algorithms proposed in this section are based on the idea that an early stopping of the algorithm, at some tree level, to go upper in the tree (see Fig. 3), may decrease the total data gathering cost, the required number of processing sensors, or both.

#### 4.1. Network Segmentation into Groups of Equal Size

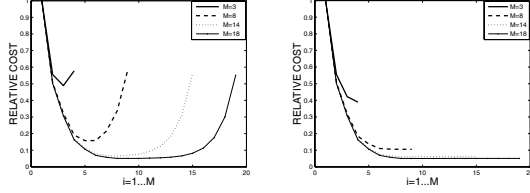
The second proposed algorithm, called *GroupNetHaar*, combines the data gathering scenario of *NetHaar* with network segmentation, with the restriction that all groups contain the same number of sensors. This provides insight into whether a spatially adaptive multiresolution approach results in total cost improvements.

Local data gathering can be performed by either (a) gathering raw data from the groups nodes to the leader-node, if only the latter has processing capabilities, or (b) with the *NetHaar* scenario, if the assumption that all nodes have processing capabilities is realistic (in other words, we consider each group as a separate sub-network and collect both LP and HP coefficients in the sensor closest to the sink). Then, the leader-node sends all transform coefficients to the network sink. Note that *NetHaar* is a special case of *GroupNetHaar*, with a network division into only one group.

#### 4.2. Adaptive Network Segmentation

With *GroupNetHaar*, the network segmentation is not automatically adapted to each specific measured signal, as the total cost corresponding to a certain group size, and thus the optimal group size, cannot be computed a-priori without global information about the measured signal. Thus, we propose an algorithm, called *AdaptNetHaar*, which provides an optimized network division according to the signal characteristics. This algorithm is fully decentralized and adapts its behavior to each individual measured signal.

The algorithm starts from the bottom of the binary tree in Fig. 3. As in *NetHaar*, the odd index sensors send their data to the even index sensors (arrows in Fig. 3). Each receiving sensor makes a decision about whether creating a 2-sensor group or keeping the present 1-sensor group state, by comparing the costs of data gathering for the new group with the sum of costs corresponding to the existing 1-sensor groups. After this operation, sensors 2, 4,  $\dots$ ,  $N$  have information about their neighbors with odd index. In the second step, they can transmit data further to the



**Fig. 4.** Relative cost versus number of groups  $n_g = 2^i$  in the network segmentation for *GroupHaar* with raw data gathering (left) and local processing inside the group (right).

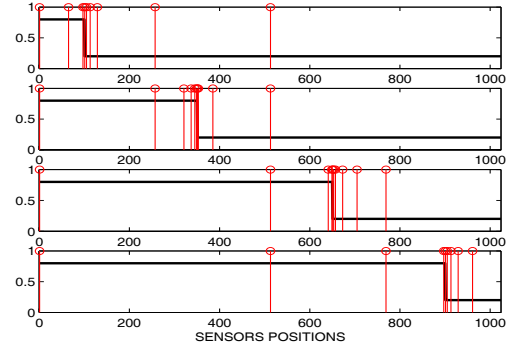
leader-nodes of 4-sensor groups (see Fig. 3). A new set of coefficients is computed at the leader-nodes (precisely, two new coefficients) and the new computed cost is compared to the cost of the previous step. The decision about collapsing the groups into one larger group is essentially determined by the values of the computed HP coefficients. The algorithm proceeds similarly for the other tree levels. For  $k = 3, \dots, M + 1$  there are  $\frac{N}{2^{k-1}}$  groups of size  $2^{k-1}$ . For each group the total cost of data gathering is compared with the gathering cost for that group corresponding to the previous algorithm steps. Further grouping is accepted only if it decreases the cost. The algorithm stops once all data reach the network sink. Note that inter-group communication in the steps of this algorithm does not incur any additional cost with respect to *NetHaar*, as communication is always done from the extremity of the network to the sink.

## 5. EXPERIMENTAL RESULTS

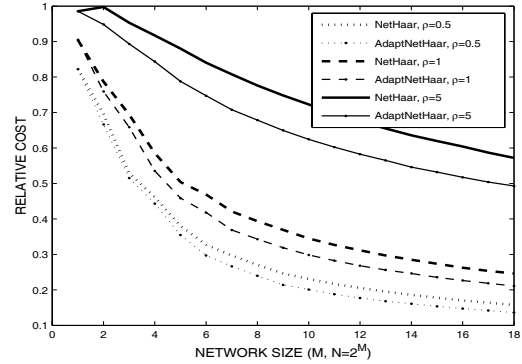
In this section we show numerical simulations for the proposed algorithms. As test signals we use piecewise constant signals with amplitude values uniformly distributed in  $[-1, 1]$ . The number of signal discontinuities varies with the network size, and it is given as  $\lceil \rho \frac{N}{\log_2 N} \rceil$ , with  $\rho$  a small real number  $\rho \in [0.1, 5]$ . Note that such signals result on average in a small number of non-zero HP coefficients compared to the total number of coefficients. For each number of discontinuities that we considered, the performance index was obtained by averaging over 1000 simulations.

Fig.4 shows the optimal fixed group size of the network segmentation with *GroupNetHaar*. Note that there is an optimal group size that minimizes the total cost. Fig.5 presents some examples of network division with *AdaptNetHaar* for a particular signals. We consider a signal with a single discontinuity that is successively placed in four different positions: at the extremity of the network, two intermediate positions, and close to the sink. The signal discontinuity position essentially determines the network segmentation: intuitively this algorithm tends to create small groups around discontinuities and larger groups of sensors on the constant pieces of signal as the measuring sensors get far from the sink.

In Fig.6, we plot the relative performance of our algorithms as a function of network size and number of discontinuities. Note that in both *NetHaar* and *AdaptNetHaar*, the decrease in relative costs increases with the network size. Also, as the number of discontinuities increases, the adaptive procedure of *AdaptNetHaar* results in significant gains over *NetHaar*. Note also that when the number of discontinuities is reduced (the level of uncertainty in the signal decreases), the gains over raw data gathering are increased. We see how the adaptive algorithm outperforms the fixed group size version for strongly varying signals (large  $\rho$ ), since adaptation to the local signal characteristic is crucial when the number of disconti-



**Fig. 5.** Network segmentation for various positions of a single discontinuity (the sink is at the extreme right).



**Fig. 6.** Comparison among the algorithms for variable number of discontinuities.

nities is large.

## 6. CONCLUSIONS AND FURTHER WORK

We studied a set of algorithms that use multiresolution wavelet methods to improve the cost of total transmission for data gathering in sensor networks that measure piecewise constant signals. We showed how important improvements can be obtained by employing local group data processing before transmission to the sink, and how even larger gains can be obtained by the use of an algorithm that adapts to the local characteristics of the signal. Our current work is focused on the study of two-dimensional scenarios.

## 7. REFERENCES

- [1] R. Cristescu, B. Beferull-Lozano, M. Vetterli: *On network correlated data gathering*, in Proc. INFOCOM'04.
- [2] A. Ciancio, A. Ortega, *A distributed wavelet compression algorithm for wireless sensor networks using lifting*, in Proc. ICASSP'04.
- [3] R. Gray: *Quantization*, IEEE Trans. Inf. Th., 1998.
- [4] D. Marco, E. Duarte-Melo, M. Liu, D. L. Neuhoff: *On the many-to-one transport capacity of a dense wireless sensor network and the compressibility of its data*, in Proc. IPSN'03.
- [5] G. J. Pottie, W. J. Kaiser, *Wireless integrated sensor networks*, Communications of ACM 2000.
- [6] A. Scaglione, S. D. Servetto: *On the interdependence of routing and data compression in multi-hop sensor networks*, MobiCom'02.
- [7] J. Acimovic, B. Beferull-Lozano, R. Cristescu, *Efficient Distributed Decorrelating Processing for Data Gathering in Sensor Networks*, to be submitted to IEEE Trans. on Signal Processing.