Performance Analysis of Distributed Source Coding and Packet Aggregation in Wireless Sensor Networks

§*L. Di Paolo*, †*C. Fischione*, §*F. Graziosi*, §*F. Santucci*, §*S. Tennina*

§*University of L'Aquila Center of Excellence DEWS*, †*Royal Institute of Technology - KTH*

e-mail: †*carlofi@ee.kth.se*, §*graziosi,santucci,stennina@ing.univaq.it*

Abstract— **In this paper, we propose a theoretical setup for evaluation of energy efficiency of wireless sensor networks (WSNs) with distributed source coding (DSC) algorithms and packet aggregation (PA). We consider four topologies for DSC and three alternatives for PA, and the system model includes a realistic network architecture with multi-hop communication, automatic repeat request protocol (ARQ), and packet losses. The analysis is carried out in two steps. Firstly we derive the packet loss probability, and then evaluate the average number of packets transmitted throughout the network. This second performance index can then be mapped onto an energy efficiency indicator. The proposed model is specifically adopted for performance comparison of the different coding strategies and aggregation schemes in terms of energy efficiency. Numerical results show that packet overheads have a relevant influence on performance, while the ARQ protocol introduces negligible effects on the energy consumption. Furthermore, DSC topologies with master-slave approach and fragmentation of packets exhibit better performance.**

Keywords: Wireless Sensor Networks (WSNs), Distributed Source Coding (DSC), Packet Aggregation (PA).

I. Introduction

Wireless sensor networks are often deployed with energy constrained nodes, wherein recharging and maintenance is not possible. Hence, the reduction of energy consumption is one of the major constraints to embody in the design of protocols and algorithms. Distributed source coding (see e.g. [1]- [2]) is a promising technique for redundancy reduction in the information flow generated by each node, thus enabling consistent energy savings. In fact, since spatial and temporal characteristics of the measured data usually exhibit significant correlation, DSC exploits the correlation structure of measurements taken by different sensors to compress data. However, DSC is only one of the techniques employed for the reduction of energy consumption, and acts above the protocol stack of a typical sensor network platform. Its interaction with lower layers of the stack (network, data link, and physical) deserves particular attention to fully exploit the advantages that DSC techniques promise. Moreover, DSC may impose some constraints to the lower layers, as for instance a suitable packet loss probability. At the same time, relevant characteristics of lower layers (e.g. packet dimension, ARQ behavior, etc.) should be accounted for when designing DSC schemes. DSC is also strongly related to the coding topology. In fact, DSC techniques show different performance according to the position of the sink node, the gathering nodes, and the field correlation structure of the data. In this context, cross-layer approaches seem suitable to design DSC scheme jointly with other traffic reduction methods.

A basic remark related to deployment of DSC in realistic environments is that header overheads in assembled data units may significantly reduce the advantages promised by DSC. Indeed, in WSNs multi-hop routing is implemented, and each node acts as a router that relays packets received from other nodes. At the same time, a node acts as a source of information that needs to transmit locally collected data. As proposed in [3], a node could aggregate packets coming from other nodes with its own data in order to reduce the impact of packet overhead, thus decreasing the generated traffic. We believe that in WSNs with multi-hop communication, DSC topologies and packet aggregation (PA) should be simultaneously designed under realistic assumptions.

Some recent contributions can be found in literature that deal separately with either DSC or PA. In [4], four topologies for DSC have been investigated with respect to packet losses and energy efficiency. In [5], the authors have studied the problem of finding the optimal aggregation alternative for a WSN that implements DSC. In [6], the problem of optimizing data gathering according to the data source correlation is investigated. Sensor clustering with routing, DSC and data source correlation was investigated in [7]. The idea of PA was first introduced in [3], where the authors proposed a framework for routing, data compression, and packet aggregation. In [8] and [9], classifications of aggregation protocols for periodic measures have been proposed.

Although the relevant existing contributions (e.g. [3]–[9]) have the merit to include detailed network scenarios in DSC, they neglect some important aspects, namely PA and fragmentation, and overhead reduction. However, the aggregation scheme without packet fragmentation is prohibitive for real multi-hop sensor networks. Consequently, the main contribution of this paper consists in joint performance analysis of DSC topologies and PA with fragmentation schemes. Specifically, we consider four alternatives of coding topology as proposed in [4], and their integration with three alternatives of aggregation techniques. Performance analysis of DSC and PA are derived in terms of the packet loss probability and average number of transmitted bytes. The last measure is taken as an

Work done in the framework of the HYCON Network of Exellence, contract number FP6-IST-511368. The work by C. Fischione was partially funded also by the Swedish Foundation for Strategic Research.

The topics of this work were partially published at IEEE GLOBECOM 2006 and ICTSM 2006

indicator of the energy efficiency. We include accurate models of the physical layer, whereas a common simplification in literature is related to the packet loss probability, which is assumed constant (as e.g. in [4]). However, the packet loss probability is dependent on the distance from the nodes and, in general, on the packet dimension. Another relevant aspect we take into account is the presence of an ARQ protocol, which increase the probability of successful packet delivery, but also the energy consumption.

The remainder of the paper is organized as follows: in Section II, the system model is described; in Section III the packet loss probability is characterized; in Section IV the average number of packets transmitted by the network is derived; in Section V, numerical results for performance evaluation are provided; finally, in Section VI, conclusions and future perspectives are given.

II. System Model

Let us consider an area over which N nodes are uniformly deployed and transmit towards a sink that acts as a gathering node. As assumed e.g in [4], the network topology is organized in a tree, where nodes belong to a level i, being i the distance of a node from the sink, in terms of number of hops (see Fig. 1). Let us denote the number of nodes at the level i with l_i, so obviously holds that $\sum_{i=1}^{L} l_i = N$. Each node is connected to a parent node and produces n_{Mis} measurements per snapshot. Each node performs the following tasks: it encodes measured

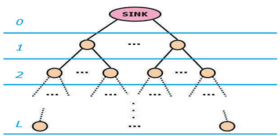

Fig. 1. Network topology with L levels.

data according to a DSC algorithm, receives packets from the child nodes and then transmits packets towards the parent node.

Throughout the paper, the packet frame specified for the Berkeley motes [10] [11] will be adopted. Therefore, a packet is composed as follows: a preamble and a synchronization symbol of P bits; a data link header of O bits, a network header of Q bits; a payload of pay_i bits that incorporates measurements performed by the generic node of the level i, and where the payload may have variable length according to the DSC topology and aggregation mechanism; CRC bits for the forward error detection; ID bits for node identification. A Manchester coding technique is employed.

A. DSC Topology

Let us denote with $Y_{i,j}$, for $i = 1 \dots L$, and for $j = 1 \dots l_i$, the data measured by the node j of the level i. We assume a static environment, i.e., the correlation properties of the data source do not change. As studied in [6], we consider four DSC coding schemes: No-Slepian-Wolf Scheme (NOSW), Sequential Slepian-Wolf scheme (SEQ), Slepian-Wolf Clustered (CL), and Slepian-Wolf Master Slave (MS).

In the NOSW topology, each sensor encodes the measures independently from the other $N-1$ nodes, and we assume that the number of bits for coding is given by the entropy of the source (see, e.g., [6]) which we denote with $H(Y_{ij})$, i.e each sensor does not perform any coding.

In the SEQ scheme, node $1,1$ performs encoding of $Y_{1,1}$ with $H(Y_{1,1})$ bits; node $1,2$ encodes $Y_{1,2}$ with $H(Y_{1,2}|Y_{1,1})$ bits under the assumption that the decoder knows $Y_{1,1}$; the node $1,j$ encodes $Y_{1,j}$ with $H(Y_{1,j}|Y_{1,1}\dots Y_{1,j-1})$ bits under the assumption that the decoder knows $Y_{1,1}\dots Y_{1,j-1}$; in general, the node i,j encodes $Y_{i,j}$ with $H(Y_{i,j}|Y_{1,1}\dots Y_{1,l_1}Y_{2,1}\dots Y_{2,l_2}\dots Y_{i,1}\dots Y_{i,j-1})$ bits, under the assumption that it knows $Y_{1,1}\dots Y_{1,l_1}Y_{2,1}\dots Y_{2,l_2}\dots Y_{i,1}\dots Y_{i,j-1}$. According to the SEQ scheme, the nodes being at a larger distance from the sink encode data with respect to those closer to the sink. This can be obviously explained considering that the packets generated by the nodes closer to the sink are more likely to be received then those farther.

In the CL case, nodes are grouped in $K = l_1$ clusters, where each cluster consists of a sub-tree having as root a node of level 1. Hence, each cluster includes the same number of nodes, which is given by N/K. For each node of a cluster, a SEQ coding topology is adopted independently from the other clusters and starting from the root node $1,k$, for $k = 1 \dots l_1$, of the clusters k.

In the MS, nodes are grouped in clusters as in the case of CL, but each root $1,k$, for $k = 1 \dots l_1$, of a sub-tree is elected as master node. Moreover, for each node of a cluster k, the DSC is performed only with respect to the master $1,k$ of the cluster the node belongs to, and not with respect to other nodes.

B. Packet Aggregation

We consider three alternatives for aggregation: the classic multihop (CMH), where the nodes relay packets without any aggregation; the aggregated multihop (AMH) , where the nodes collect the received packets, aggregate them at the MAC layer into a single frame and forward it; the Fragmented Aggregated Multihop (FAMH) , which is similar to AMH but with the difference that, when an aggregated packet reaches a dimension larger than a maximum threshold, it is fragmented in multiple packets.

C. Performance index

We consider as a performance parameter the average number of bits per snapshot transmitted by the network and we denote this by E_N. In the computation of E_N, we will neglect the impact of acknowledgement packets, which is a rather common assumption in literature, and is motivated by the fact that they have marginal impact on overall performance. Nevertheless, computation of E_N requires the characterization of packet error probability under different assumptions for DSC and PA. We approach this task in the next Section.

III. Performance Analysis: Packet Error Probability

We consider a wireless channel which exhibits non-selective behavior both in frequency and in time, and a binary FSK modulation scheme. These assumptions are representative of actual transceivers working in the ISM frequency band and are basically the same assumptions adopted e.g. in [12]. Therefore, the bit error probability is expressed as follows [13]:

$$P_e(d) = \frac{1}{2+\gamma_0(d)} \ , \quad (1)$$

where $\gamma_0(d)$ is the average value of the Signal to Noise Ratio (SNR) computed at the distance d from the signal source. We adopt the following model for $\gamma_0(d)$ in dB [14]:

$$\gamma_0(d)_{dB} = P_{t\ dBm} - P(d)_{dB} - P_{n\ dBm} \ , \quad (2)$$

where P_t is the transmitted power, $P(d)$ is the path loss at the distance d from the transmitter, and P_n is the noise floor at the receiver. The path loss is expressed as follows:

$$P(d) = P(d_0) + 10\alpha \log_{10}(\frac{d}{d_0}) \ , \quad (3)$$

where $P(d_0)$ is the path loss computed at the reference distance d_0 (see e.g. [15]); α is the path-loss decay constant, and lays in the interval $(2\dots4)$. The noise floor is assumed to be expressed as $P_n = (F+1)kT_0B$, where F is the noise figure of the receiver, k is the Boltzmann constant, T_0 is the room temperature and B is the equivalent noise bandwidth of the receiver.

Under the assumptions that the CRC code is always able to detect perfectly erroneous packets (see e.g. [16] for an experimental support to this assumption), the packet loss probability turns out to be expressed as follows:

$$P_{SH}(d, pay_i) \quad = \quad 1-[1-P_e(d)]^{2(P+O+Q+CRC+pay_i)} \quad (4)$$

where pay_i is the average payload length from the nodes of level i.

We define the average packet loss probability along one hop of the level i as the average of (4) with respect to the hop length:

$$P_{SH}(pay_i) = \int_{d_{min}}^{d_{max}} \frac{P_{SH}(d, pay_i)}{d_{max}-d_{min}} dd \ , \quad (5)$$

where d_{min} and d_{max} are, respectively, the minimum and maximum possible distances between any pair of nodes child-parent.

When an ARQ protocol is implemented at the level i, with M_i maximum retransmissions, we can easily express the packet loss probability as follows:

$$P_{SH}(pay_i, M_i) = P_{SH}(pay_i)^{M_i} \ . \quad (6)$$

We evidence that in (6) the dependence of the multi-hop packet loss probability on the aggregation scheme is included in $P_{SH}(pay_i, M_j)$, which will be characterized in the next subsections for the cases of aggregation scheme CMH, AMH, and FAMH.

Some definitions will be useful to study the packet loss probability. We define the average connectivity C_i of a node of the level i as the average number of children of the node, and is expressed as

$$C_i = \frac{l_{i+1}}{l_i} \quad i = 0, \dots, L-1 \ , \quad (7)$$

Note that $l_0 = 1$ is the number of nodes of the level 0 (only the sink), while $C_L = 0$.

A. CMH

The payload length for the CMH scheme at the level i, for readings generated by a node of the level j, can be expressed as follows:

$$pay_i = n_{Mis}H_j \qquad i = 1, \dots, L \text{ and } j \geq i \ , \quad (8)$$

where H_j is the coding rate at the level j as it is determined by the one of DSC topology.

The single-hop packet loss probability is therefore given by (5), plugging into (8).

B. AMH

In the AMH scenario, the following relations hold true:

$$pay_i = \begin{cases} ID + n_{Mis}(H_i + \\ \quad \sum_{j=1}^{L-i} H_{i+j} \prod_{k=1}^{j} n_{i+k}) & i = 1, \dots, L-1 \\ ID + n_{Mis}H_L & i = L \ , \end{cases} \quad (9)$$

$$n_i = [1 - P_{SH}(pay_i, M_i)]C_{i-1}, \quad i = 2, \dots, L \ , \quad (10)$$

where n_i is the average number of packets of the level i that are successfully received by the parent node of the level $i-1$.

In order to prove (9) and (10), let us first consider the average number of packets that are successfully transmitted to each parent-node of the level $L-1$:

$$n_L = [1 - P_{SH}(pay_L, M_L)]C_{L-1} \ . \quad (11)$$

At the level $L-1$, each node aggregates the received packets and relays a unique packet having a payload as follows:

$$\begin{aligned} pay_{L-1} \quad &= \quad ID + n_{Mis}H_{L-1} + pay_L n_L \\ &= \quad ID + n_{Mis}(H_{L-1} + H_L n_L) \ . \quad (12) \end{aligned}$$

At the level $L-2$, the average number of received packets is $n_{L-1} = [1-P_{SH}(pay_{L-1}, M_{L-1})]C_{L-2}$. Hence, the payload and average number of packets coming to such a level are:

$$\begin{aligned} pay_{L-2} \quad &= \quad ID + n_{Mis}(H_{L-2} + H_{L-1}n_{L-1} + \\ &\qquad H_L n_L n_{L-1}) \ , \quad (13) \end{aligned}$$

$$n_{L-2} = [1 - P_{SH}(pay_{L-2}, M_{L-2})]C_{L-3} \ . \quad (14)$$

The generalization of (13) and (14) to obtain (9) and (10) is straightforward.

Finally, the single-hop packet loss probability is given by (5), using (9).

C. FAMH

In the FAMH case, the payload dimension is upper bounded by a maximum dimension MAX_{PAY}. When an aggregated packet would exceed this bound, it is split into F_i fragments, with $F_i = \lceil \frac{pay_i}{MAX_{PAY}} \rceil$, where $\lceil \cdot \rceil$ denotes the upper integer (ceiling). Each fragment will carry a payload of the level 2 protocol stack whose length is

$$payF_i = (Q + pay_i)/F_i \ . \quad (15)$$

The fragmented packet loss probability averaged over a single hop is given as follows:

$$P_{SH\ F_i}(payF_i) = \int_{d_{min}}^{d_{max}} \frac{P_{SH}(d, payF_i)}{d_{max}-d_{min}} dd \ , \quad (16)$$

where

$$P_{SH\ F_i}(d, payF_i) \quad = \quad 1 - [1 - P_e(d)]^{2(P+M+CRC+payF_i)} \quad . \tag{17}$$

The packet loss probability over a single hop can be expressed as

$$P_{SH}(pay_i) = 1 - [1 - P_{SH\ F_i}(payF_i)]^{F_i} \quad , \tag{18}$$

Finally, (18) can be used in (6) replacing $P_{SH}(pay_i)$ to obtain the single-hop packet loss probability.

IV. Performance Analysis: Average Number of Transmitted Bits

The derivation of the average number of bits transmitted by the networks requires the computation of the average number of transmitted packets at the generic level i. With this goal in mind, let us denote with $desc_i$ the average number of descendants of the generic node of the level i whose packets are successfully received at that node, whereas let us indicate with $desc_{i,j}$ the average number of descendants at the level j whose packets are successfully received at the generic node of the level i (obviously, $i < j$). In the following subsections, we derive $desc_i$ and, thus, E_N for the CMH, AMH and FAMH scenarios.

A. CMH

At the level i of the network, the following expression holds true:

$$desc_i = \sum_{j=i+1}^{L} desc_{i,j} \quad , \qquad i = 0, \ldots, (L-1) \quad , \tag{19}$$

where:

$$desc_{i,j} = \prod_{k=i+1}^{j} [1 - P_{SH}(pay_j, M_k)] C_{k-1} \quad . \tag{20}$$

We can proof the validity of (19) and (20) as follows. At the level $L-1$, $desc_{L-1}$ is easily obtained by the subsequent expression:

$$desc_{L-1} = desc_{L-1,L} = [1 - P_{SH}(pay_L, M_L)] C_{L-1} \quad . \tag{21}$$

At level $L-2$, $desc_{L-2}$ depends on the number of packets that are generated from level L and level $L-1$ (where the number of nodes of such levels are obviously taken into account) and are correctly received at level $L-2$:

$$\begin{aligned} desc_{L-2} &= desc_{L-2,L-1} + desc_{L-2,L} = \\ &= [1 - P_{SH}(pay_{L-1}, M_{L-1})] C_{L-2} + \\ &\quad [1 - P_{SH}(pay_L, M_{L-1})] \cdot \\ &\quad [1 - P_{SH}(pay_L, M_L)] C_{L-2} C_{L-1} \quad . \end{aligned} \tag{22}$$

For the generic level i, the expression (22) can be extended in a straightforward way.

The number of retransmissions of a packet sent from the node of level i over the hop between the level i and $i-1$ is denoted as $TxARQ_i$, and can be expressed as

$$TxARQ_i = \frac{1 + M_i P_{SH}(pay_i)^{M_i+1} - (M_i+1) P_{SH}(pay_i)^{M_i}}{1 - P_{SH}(pay_i)} . \tag{23}$$

Therefore, the average number of transmitted bits from the generic node of level i, when an ARQ protocol is employed, is given by the following expression:

$$\begin{aligned} B_{TX\ i} \quad &= \quad (P + O + Q + pay_i + CRC)\, TxARQ_i + \\ &\quad \sum_{j=i+1}^{L} (P + O + Q + pay_j + CRC) \cdot \\ &\quad desc_{i,j} TxARQ_j \quad . \end{aligned} \tag{24}$$

Finally, the average number of transmitted bits is:

$$E_N = \sum_{i=1}^{L} B_{TX\ i} l_i \quad . \tag{25}$$

B. AMH

In this scenario, each node of the level i transmits only one packet containing both the measurements taken from local sensing and the measurements coming from lower levels and encapsulated therein. Therefore, E_N is readily expressed as follows:

$$E_N = \sum_{i=1}^{L} (P + O + Q + pay_i + CRC)\, TxARQ_i l_i \quad . \tag{26}$$

C. FAMH

In the fragmented aggregated multi-hop scenario, it is easy to see that the average number of transmitted bits is expressed as

$$E_N = \sum_{i=1}^{L} (P + O + payF_i + CRC) \;\; F_i\, TxARQ_i \;\, l_i \quad , \tag{27}$$

where $TxARQ_i$ is computed using $payF_i$ defined in (15).

V. Numerical Results

We consider a network deployed over a squared area having facet ℓ. The area is divided in $S \times S$ equal dimension sub-squares, having facet ℓ/S. The sink node is located in the middle of the area. We include in the network $N = 64$ nodes, thus $S = 8$, where a node is located according to a uniform distribution within each sub-square. To avoid a superposition among sensors of adjacent sub-squares, each node is located inside a cell centered in the sub-square, but having dimension $\varepsilon\ell/N$ with $\varepsilon < 1$. The association node-level is based on the position of the node inside the squared area. Those nodes that are located in the first squared tier around the sink, belong to level 1. The nodes located in the second squared tier around the sink belong to level 2, and so on. Therefore, there are $L = 4$ levels in the network. Moreover, each node is supposed to sense $n_{Mis} = 8$ measurements per snapshot.

The values of d_{min} and d_{max} for the not-border cells can be computed as follows (see Fig. 2): $d_{max} = \sqrt{2}(1+\varepsilon)\frac{\ell}{N}$ $d_{min} = (1-\varepsilon)\frac{\ell}{N}$, while for the nodes across the borders, they are $d_{max} = \frac{1+\varepsilon}{\sqrt{2}}\frac{\ell}{N}$ and $d_{min} = \frac{1-\varepsilon}{\sqrt{2}}\frac{\ell}{N}$. The ε factor is set to 0.9. d_{max} is set to $7m$ and hence $\ell = 20.84m$. The values set for the number of nodes for each level and average connectivity are reported in Tab. I, whereas the values of the packet format are reported in Tab. II.

The transmission power has been set to $5dBm$. The receiver is assumed to work at a room temperature of 300 K, with a noise figure of $13dB$, an equivalent bandwidth of 30 KHz. Hence, the noise floor is $P_n = -115\ dBm$. Without loss of generality, we set the maximum retransmission iterations of the ARQ protocol for the levels $1, ..., L$, respectively, as $3, 3, 1, 1$.

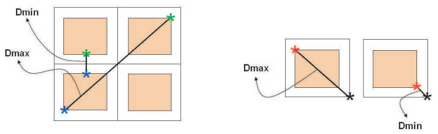

Fig. 2. d_{min} and d_{max} for the nodes of level $i < L$ (left) and $i = L$ (right).

We compute the entropy of the source by modelling the spatial correlation among sensed data according to a Gaussian random field (see e.g. [6] and [17]). Adopting this model, the measures are characterized with a N dimensional multi-variate normal distribution $G_N(\boldsymbol{\mu}, \mathbf{K})$ having average $\boldsymbol{\mu}$ and covariance matrix $\mathbf{K} = [K_{i,j}]$, where $K_{ii} = \sigma_i$ and $K_{i,j} = \sigma_i^2 e^{-ad_{ij}}$ for $i \neq j$, where a is the spatial correlation coefficient, and d_{ij} is the distance between the nodes i and j. Note that the correlation coefficient tends to 1 for highly correlated sources, whereas tends to zero for scarcely correlated sources. The entropies H_i for each level in the network can thus be computed by resorting to the well known conditional entropy of Gaussian sources (see e.g. [6] and [18]). In Tab. III, the values of the entropies are reported for each level and coding topology, with $\sigma_i = a = 1$. It is relevant to observe that the SEQ topology exhibits the lower values of entropies, while the MS a larger one. Consequently, the SEQ topology seems the most attractive in terms of bit and energy consumption.

i	l_i	C_i
1	4	4
2	12	3
3	20	$20/12$
4	28	$28/20$

TABLE I

NUMBER OF NODES FOR EACH LEVEL AND AVERAGE CONNECTIVITY.

P	O	Q	MAX_{PAY}	ID	CRC
80	40	56	176	8	16

TABLE II

PACKET FORMAT.

	H_1	H_2	H_3	H_4
SEQ	5.7833	4.1389	4.0167	4.0238
CL	10	4.4778	4.1667	4.1333
MS	10	4.75	5	5.2810

TABLE III

AVERAGE ENTORPY PER LEVEL.

$\uparrow E_N(ARQ)$	CMH	AMH	$FAMH$
$NO\ SW$	203.3054	165.4234	225.5351
SEQ	141.8292	64.1911	66.6674
CL	143.5914	66.9636	68.9328
MS	151.8241	79.3944	148.0798

TABLE IV

INCREASING OF E_N AFTER THE ARQ PROTOCOL.

In Fig. 3, the average number of bytes transmitted by the generic node of level i is reported. Each level is subject to a different traffic load, especially in the case of CMH. In the presence of aggregation schemes, the difference between the first level and the last level in terms of transmitted bytes is seen to decrease. This behavior results in evident advantages for the nodes closer to the sink, which are typically exposed to faster energy depletion.

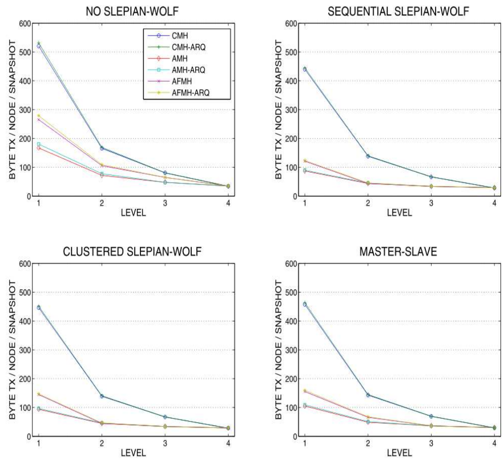

Fig. 3. Average number of byte, $B_{TX\ i}$, transmitted per node of the level i.

In Fig. 4, the values of E_N for the different DSC topologies and aggregation algorithms are reported. The right column is referred to the ARQ protocol, while the left column is referred to the case where ARQ is not adopted. From Fig. 4, it is evident that CMH exhibits the worst performance, whatever DSC scheme is adopted. Similarly, the NOSW topology exhibits the worst performance, while better performance can be achieved, respectively, with CL, MS, and SEQ. In the last three cases, there are not relevant differences, and this can be mainly due to the large spatial correlation. By the contrary, aggregation algorithms provide significant performance improvement. The difference between AMH and FAMH is more evident in the NOSW topology, where a number of packets transmitted over the network are observed.

In Tab. IV, the increase of E_N that is induced by the ARQ protocol is reported for the different aggregation schemes. Although ARQ introduces a larger number of packets per node to be relayed, we can observe that it leads to negligible decreasing of performance in terms of percentage values of E_N.

For a better understanding of E_N behavior, let us define the

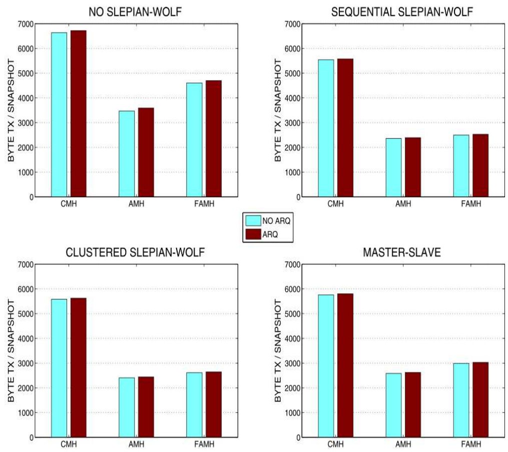

Fig. 4. Average number of byte transmitted by the network. The right bars are referred to ARQ case, whereas the left ones are referred to the case without ARQ.

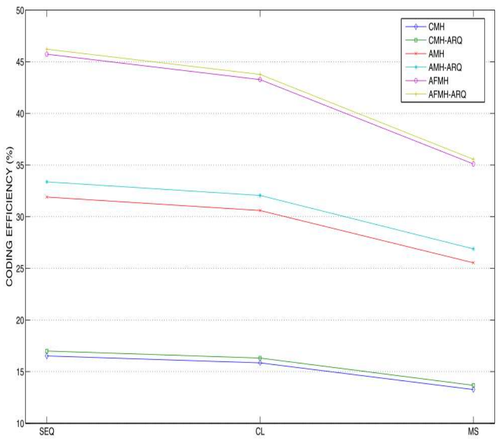

Fig. 5. Coding Efficiency for CMH, AMH, and FAMH with and without ARQ. On the abscissa are reported the cases SEQ, CL, and MS.

following coding efficiency metric:

$$\eta_{CT} = 1 - \frac{E_{N,CT}}{E_{N,NOSW}} \ , \quad (28)$$

where the subscript CT denotes one of the coding topologies SEQ, CL, or MS. In Fig. V, the coding efficiency is reported for the different coding topologies (x axis) and aggregation algorithm. The coding efficiency increases remarkably when we move from CMH to FAMH, with AMH providing intermediate performance. Once again, it is evident that efficiency is slightly improved by the use of ARQ protocol. Finally, we can observe that the performance provided by the three topologies SEQ, CL and MS are slightly different for the cases of CMH and AMH, while more they are more evident in FAMH.

VI. CONCLUSIONS AND FUTURE PERSPECTIVES

In this paper, performance analysis of distributed source coding and packet aggregation have been considered for a multi-hop single sink wireless sensor network. Specifically, we provided expressions for coding efficiency and packet loss for three cases of packet aggregation mechanisms (classic multi-hop, aggregate multi-hop and aggregated and fragmented multi-hop) and four cases of distributed source coding topology (no Slepian-Wolf coding, sequential, cluster and master-slave).

Numerical results show that packet overheads and packet loss probability may play a significant role in network energy consumption. Better performance in term of efficiency are obtained with the sequential coding topology and fragmented aggregated mechanism. The lowest absolute energy consumption is obtained with the sequential scheme with aggregation without fragmentation. Finally, the introduction of an ARQ protocol, while decreasing the probability of discarding packets, does not affect significantly the energy consumption index.

Besides the results presented in this paper, further work is in progress and is focused on the characterization of performance in term of multi-hop packet loss probability. We are also interested on development of our framework in order to include further cross-layer constraints, e.g. effects of time-delay sensitive applications, time variant wireless propagation scenario, and non stationary data source. Furthermore, the analytical framework addressed in this paper will be developed to include cross layer optimization of clustering, coding and power control.

REFERENCES

[1] S.S. Pradhan, K. Ramchandran. Distributed source coding: Symmetric rates and application to sensor networks. *Proceedings of the Data Compression Conference (DCC)*, 2000.
[2] J. Chou, D. Petrovic, K. Ramchandran. A distributed and adaptive signal processing approach to reducing energy consumption in sensor networks. *IEEE INFOCOM*, 2003.
[3] D. Petrovic, R. C. Shah, K. Ramchandran, J. Rabaey. Data funneling: Routing with aggregation and compression for wireless sensor networks. *IEEE SNPA*, 2003.
[4] D. Marco, D. L. Neuhoff. Reliabiity vs. efficiency in distributed source coding for field-gathering sensor networks. *IEEE IPSN*, 2004.
[5] Sueng Jun Baek, G. de Veciana, Xun Su. Minimizing energy consumption in large-scale sensor networks through distributed data compression and hierarchical aggregation. *IEEE Journal on Selected Areas in Communications*, 2004.
[6] R. Cristescu, B. Beferull-Lozano, M. Vetterli. On network correlated data gathering. *IEEE INFOCOM*, 2004.
[7] S. Pattem, B. Krishnamachari, R. Govindan. The impact of spatial correlation on routing with compression in wireless sensor networks. *IEEE IPSN*, 2004.
[8] I. Solis, K. Obraczka. The impact of timing in data aggregation for sensor networks. *IEEE ICC*, 2004.
[9] U. Roedig, A. Barroso, C. J. Sreenan. Determination of aggregation points in wireless sensor networks. *IEEE EUROMICRO*, 2004.
[10] *CrossBow Technologies*. http://www.xbow.com.
[11] *TinyOS*. http://webs.cs.berkeley.edu/tos/.
[12] H. Karvonen, Z. Shelby, C. Pomalaza-Rez. Coding for energy efficient wireless embedded networks. *IWWAN*, 2004.
[13] S. Haykin. *Communication System*. Wiley, 2001.
[14] M. Zuniga, B. Krishnamachari. Analyzing the transitional region in low power wireless links. *SECON*, 2004.
[15] K. Soharabi, B. Manriquez, G. Pottie. Near ground wideband channel measurement. *IEEE Vehicular Technology Conference*, 1999.
[16] J. Jeong, C. Tien Ee. Forward error correction in sensor networks. Technical report, Department of Electrical Engineering and Computer Science, University of California, Berkeley., 2003.
[17] D. Ganesan, R. Cristescu, B. Beferull-Lozano. Power-efficient sensor placement and transmission structure for data gathering under distortion constraints. *IEEE IPSN*, 2004.
[18] T. M. Cover, J. A. Thomas. *Elements of Information Theory*. Wiley, 1991.