A novel trade-off between communication and computation costs for data aggregation in wireless sensor networks

N. Arastouie*, M. Sabaei, V. Hakami and S. Soltanali

Computer Engineering and Information Technology Department,
Amirkabir University of Technology,
PO. Box 15875-4413,
424 Hafez Avenue, Tehran, Iran
Fax: +98 (21) 66419728
E-mail: arastouie@aut.ac.ir
E-mail: sabaei@aut.ac.ir
E-mail: vhakami@aut.ac.ir
E-mail: Ssoltanali@aut.ac.ir
*Corresponding author

Abstract: Wireless Sensor Networks (WSNs) consist of inexpensive low-power miniature sensing devices with severe power constraints, necessitating energy-efficient solutions for networking operations. Major prior art proposals have been primarily directed towards minimising the communication cost, either implicitly assuming away the computation overhead as being negligible, or radically trading against it. However, in computation-bound scenarios, dealing with a large volume of data, such simplifying assumptions or radical measures tend to be inefficient. In this paper, we investigate the problem of minimising the overall energy needed to send data from a set of sensor nodes to a single destination, where each node is in charge of a mission. Two types of missions are defined: sensing and decision making; while source nodes are only in charge of sensing, relay nodes can carry out both missions simultaneously. More specifically, given a node's current backlog and its latest view on the relevant portion of the data-gathering tree, taking on the decision-making mission involves deriving an online trade-off between energy costs of compression and communication, and deciding between sending data either in the raw mode or alternatively compressed with a feasible optimal compression ratio. The used data compression technique depends on the type of application and the spatiotemporal correlation in the packets. Simulation experiments reveal that, compared with previous methods, the proposed scheme exhibits superior energy efficiency with an additional 36% reduction of the costs.

Keywords: sensor network; energy consumption; computation cost; aggregation.

Reference to this paper should be made as follows: Arastouie, N., Sabaei, M., Hakami V. and Soltanali S. (2013) 'A novel trade-off between communication and computation costs for data aggregation in wireless sensor networks', *Int. J. Ad Hoc and Ubiquitous Computing*, Vol. 12, No. 4, pp.245–253.

Biographical notes: Narges Arastouie received her BSc in Computer Engineering in 2007 and the MSc in Information Technology in 2009 both from Amirkabir University of Technology (AUT), Tehran, Iran. She is currently pursuing her PhD in Information Technology at the same University. Her research interests include computer networks, sensor networks, data correlation and aggregation.

Masoud Sabaei received his BSc in Computer Engineering from Isfahan University of Technology, Isfahan, Iran, in 1992, and his MSc and PhD from Amirkabir University of Technology (AUT), Tehran, Iran, in 1995 and 2001, respectively. He joined the Computer Engineering and Information Technology (CEIT) Department of Amirkabir University of Technology in 2004. Currently, as an assistant professor, he is the Director of the Mobile Ad hoc and Wireless Sensor Lab at AUT. His research interests include high-speed networks, network management, mobile ad hoc and sensor networks and opportunistic network.

Vesal Hakami received his BSc in Computer Engineering in 2004, and the MSc in Information Technology in 2008 both from the Amirkabir University of Technology (AUT), Tehran, Iran. He is presently pursuing his PhD in Information Technology at the same university. His current research

interests include context-aware and mobile computing middleware in ad-hoc mobile environments, self-stabilising protocols, quorum-based power-saving schemes as well as target tracking and query processing in wireless sensor networks.

Saeed Soltanali received his BSc in Computer Engineering in 2004 from the Bahoonar University, Kerman, Iran and the MSc in Information Technology in 2007 from the Amirkabir University of Technology (AUT), Tehran, Iran. He is presently pursuing his PhD in Computer Engineering at the Tehran university. His current research interests include game-theory and mobile computing middleware in ad-hoc mobile environments, tracking and query processing in wireless sensor networks.

1 Introduction

A Wireless Sensor Network (WSN) consists of sensor nodes, which are small in size and have limited sensing, processing and transmission capabilities. Several applications of these networks have been proposed, such as healthcare systems, military, home and rescue situations.

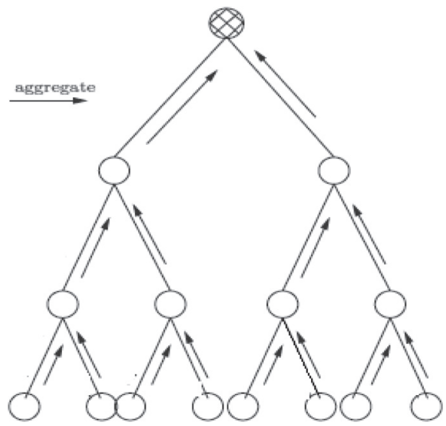
The wireless transmission of data packets consumes most of the energy in a WSN, especially when there exist multi-hop paths from source nodes towards the sink(s). Coming up with an optimal policy for energy cost management becomes more challenging when dealing with scenarios involving a huge volume of sensed data to be forwarded across the network. Transmission of video frames in an object-tracking application is an example in which the locations of the objects are needed to be reported to the sink with respect to a predefined deadline (Naderan et al., 2012). In spite of the significant improvements in compression techniques, scrupulous trade-offs still need to be made between the real-time sending of data packets and their aggregated transmission in compressed form.

Data aggregation can be performed in either of the centralised, tree-based, static cluster-based and dynamic cluster-based fashions (Fasolo et al., 2007). While the existing schemes have come up with effective ways of aggregating the sensed data, there has been little care for making realistic tradeoffs between computation and communication costs, and thus conducting in-network compressions in a tunable style.

The main idea behind this work is to cut down on the energy costs involved in delivering packets across a data-gathering tree by delegating some decision-making authority to the intermediate nodes to conduct online cost analysis and optimally decide on the specifics of the transmission of their data to the sink. Given the node's current backlog and armed with the latest view on the relevant portion of the tree, the outcome of this analysis determines whether it is more efficient to send data 'as is' or alternatively work out an optimal level of compression subject to the feasibility constraints imposed by the compression algorithm. This holistic trade-off between computation and communication costs is often a neglected aspect in the majority of the existing data aggregation methods (Leone et al., 2005; Goel and Estrin, 2005) given that their prime concern has been devoted solely to the minimisation of the communication costs.

We envision two types of missions for the nodes: sensing and decision making; the nodes only in charge of sensing are referred to as 'source'. 'Relay' nodes, on the other hand, can act in both roles concurrently. Periodically and in each communication round, the nodes appointed with a sensing mission transmit their measured data to their predecessor along the data-gathering tree, as is illustrated in Figure 1. Decision-making nodes, in turn, weigh the options of transmitting with or without compression with respect to the incurred end-to-end costs. Unlike comparable prior works (Yu et al., 2008; Eswaran et al., 2009), this cost analysis is not merely driven by the flow of a single leaf source node at the extreme downstream end of a flow path, but rather the aggregation process is governed dynamically with the backlog of all contributing sources including the current node's as well. The actual data compression is assumed to be carried out by an application-specific technique and with the consideration of the spatiotemporal correlations amongst data packets. We henceforth refer to our proposed scheme as the Adaptive approach for Reducing Energy consumption considering nodes' Mission (AREM).

Figure 1 Tree-based aggregation



The remainder of this paper is organised as follows: Section 2 gives a brief account on the background of our research and guides the reader through related work in this area. Section 3 is devoted to the discussion of our proposed scheme for energy cost management. Numerical results derived from our simulation experiments are presented in Section 4. The paper ends with a concluding epilogue in Section 5.

2 Related work

A considerable number of previous studies have attempted to reduce the energy consumption through applying in-network processing measures. Their fundamental idea is to reduce the number of packets propagated within the network with the help of techniques such as redundant data removal and merging (Fasolo et al., 2007; Leone et al., 2005).

The exploitation of tree-like structures for data gathering in WSNs has been investigated extensively in the relevant literature; for instance, Acimovic et al. (2005) have developed adaptive distributed algorithms for power-efficient data gathering in sensor networks and Liang et al. (2007) have applied online data-gathering schemes for maximising network lifetime in sensor networks. In these works, the authors assume that there is a workload of data-gathering queries, which arrive in sequence. To respond to each query as it arrives, the system builds a routing tree for it. Within the tree, the volume of the data transmitted by each internal node depends not only on the volume of the sensed data by the node itself, but also on the volume of the data received from its children. With no prior assumption on the specifics of the future query arrivals and generation rates, the objective is to maximise the number of answered data-gathering queries until the first node in the network fails.

Many proposals fall into the broad mainstream of research in the area of data aggregation schemes for WSNs (Liang et al., 2007; Zhu et al., 2005; Dong et al., 2010; Kanna and Iyengar, 2004; Hong and Prasanna, 2004; Choi and Das, 2005; Tahir and Farrell, 2009; Haque et al., 2009; Zytoune et al., 2011; AL-khdour and Baroudi, 2009). Among the more closely related works to our scheme, in Liang et al. (2007), the authors tackle with the problem of monitoring values of every sensor node over time. It can be considered as a degenerate version of an aggregation function where a node does not actually aggregate or merge its received or sensed data; instead, there is a basic notion of spatiotemporal correlation between the sensor values and a periodic monitoring scheme tries to exploit it in a similar way to an aggregation scheme. It has been argued that such spatiotemporal redundancy can also be exploited in video compression to save storage size or bandwidth usage. In particular, a central node calculates and transmits predictions to all nodes. Each node, then, sends its update only if it is different from the prediction received from the central node. The scheme proposed in Liang et al. (2007) raises the question of whether the transmission of prediction can actually be efficient in a large random network or not. Unfortunately, these algorithms have been tested only in small networks where every node is assumed to be at single-hop distance from the central node.

Our approach, with its focus on joint compression and communication, is closest in spirit to recent works presented in Yu et al. (2008) and Eswaran et al. (2009). In Yu et al. (2008), have developed a data-gathering tree algorithm with tunable compression for communication, named as the Tunable-Compression-based Data Gathering (TCDG). To tune data compression over the gathering tree, a flow-based model has been proposed where data from each source is compressed and transmitted as a flow over the corresponding path from the source to the sink. In Eswaran et al. (2009), have taken on a similar perspective, but they have also come up with a utility-based rate control model to account for the problem jointly with congestion control in a WSN consisting of multiple competing missions with different utilities.

While we share the same motivations as Yu et al. (2008) and Eswaran et al. (2009), we instead develop an adaptive approach for reducing energy consumption considering node missions in WSNs. A mission is carried out by applying compression in diverse levels after conducting a comprehensive online cost analysis at each intermediary node. We also proffer a more accurate model of compression cost when compared with Yu et al. (2008) and relax the assumption that limits the sources to the leaves of the data-gathering tree.

3 System model

In this section, we first present the network model and basic assumptions, and then go on with outlining the proposed generic models of computation and communication costs. The last subsection is devoted to the specifics of the cost analysis for our online trade-off scheme.

3.1 Network model

We assume that the sensor network features the following properties:

1. It consists of N nodes, denoted by $s_1, s_2, \dots, s_N$
2. It is a static densely deployed network over a two-dimensional geographic space of size A×A. The nodes are distributed uniformly and have no movement.
3. The positions of each node are represented by (X, Y) in a two-dimensional coordinate system, with X and Y being random variables uniformly distributed within the range [0, A].
4. The network is modelled as a connected weighted graph $G = (V, E, W)$, where the vertex set V represents the sensor nodes as well as the sink; the edge (link) set E represents the wireless connection between nodes and a weight we is associated with each edge $e \in$ E.
5. w_e is the energy cost of transmitting a data packet with unit size over e. The edge weight is determined by the distance between two adjacent nodes, the radio device and the communication environment.
6. There exists just one sink node.
7. The energy of the sensor nodes cannot be replenished, i.e., a sensor node will die if its energy is exhausted.
8. Each node is assigned a mission in the network.
9. Two types of missions are defined: sensing and decision making. A source node's mission is only sensing, while relay nodes can be relegated with both sensing and decision making.

10. Let the jth mission be denoted as mj, and M be the set of all mission $m_i \in$ M. Hence, M includes three types of missions: sensing, decision-making, and the combination of both sensing and decision-making.
11. $mj(i)$ is defined as the jth mission of the ith node, $i \in$ E.
12. Each source node generates a data packet of unit size.
13. Network nodes feature homogeneous energy capacities.
14. A data aggregation tree is a subtree of G rooted at the sink and denoted by T = $<V', E'>$, where $V' \subseteq V$ and $E' \subseteq E$.
15. A simple communication mechanism is assumed, which is based on a Medium Access Control (MAC) protocol that guarantees collision- and interference-free packet delivery.
16. For the purpose of this paper, we assume a lossless compression process using *gzip*, which is capable of supporting multiple levels of compression ratios (Barr and Asanovi, 2003).

Table 1 lists the symbols and notations used throughout the paper.

Table 1 Table of notations

The graph representing the sensor network with V as the set of sensor nodes, E as the set of links, w as the weights associated with the edges in E.	$G = <V,E,w>$
Set of source nodes	$R \subseteq V$
The sink node in V	*Sink*
The weight of edge $e \in$ E	*we*
Relative computation cost for compressing	γ
Joint entropy of $i \geq 1$ unit data	*Hi*
Data entropy rate, i.e., $\rho = H_1$	ρ
The set of missions in the network	*M*
The jth mission in the ith node ($i \in V$)	$mj(i)$
The energy consumed at node i with mission mj for transmitting message of size f_o as output	$\varepsilon_{TX^i_{mj(i)}}(f_o)$
The energy consumed at node i for receiving message of size $\acute{F}$ as input	$\varepsilon_{RCV}(\acute{F})$
The energy consumed at node i with mission mj for generating compressed data of size f_o from the uncompressed backlog of size F	$\varepsilon_{comp^i_{mj(i)}}(F, f_o)$
The energy consumed at node i for decompressing data of size fo	$\varepsilon_{decomp}(fo)$
The overall energy required for the preparation and transmission of compressed data of size fo from an uncompressed backlog of size F along the remaining portion of the path to the sink; $\varepsilon_{\mathrm{COM}}$ is estimated by each node i with mission mj, where $mj(i)$ prescribes a 'relay' role for i.	$\varepsilon_{\mathrm{comp}^i_{mj(i)}}(F, f_o)$
The overall energy required for the transmission of the raw backlog of size $\acute{F}$ along the remaining portion of the path to the sink. $\varepsilon_{\mathrm{RAW}}$ is estimated by each node i with mission mj, where $mj(i)$ prescribes a 'relay' role for i.	$\varepsilon_{\mathrm{RAW}^i_{mj(i)}}(\acute{F})$
$\Delta^i_{mj(i)} = \varepsilon_{RAW^i_{mj(i)}}(\acute{F}) - \varepsilon_{COM^i_{mj(i)}}(F, f_0)$	$\Delta^i_{mj(i)}$

3.2 *Energy cost model for the compressed-communication scenario*

From the perspective of a given relay node, the energy cost associated with the preparation and delivery of the input data in compressed form is directly related to the sum of the sizes of the input streams from its immediate children in the aggregation tree together with the length of its own information (in case it is delegated with a sensing mission as well). We denote this sum by $\acute{F}$. If the length of each stream is represented by f_i ($i \in N$, and if n denotes the total number of inputs including the node's own load, $\acute{F}$ can be calculated as:

$$\acute{F} = \sum_{i=1}^{n} f_i. \tag{1}$$

Each 'input' size f_i can be attributed to one of the two types of packets: compressed or uncompressed; in effect, $\acute{F}$ represents the total length of inputs while each input, in turn, may consist of a compressed or uncompressed packet; we distinguish between the inputs received from all children by the following labels:

$$\begin{cases} \beta_i & \text{the input received from child } i \text{ is in compressed form} \\ S_i & \text{the input received from child } i \text{ is in uncompressed form} \end{cases}. \tag{2}$$

The cost analysis for each round of a decision-making mission should factor in the overhead associated with the decompression of all inputs of type β so as to be able to efficiently aggregate the input flows with respect to the spatiotemporal correlations amongst the data packets. Once the entire backlog is available in uncompressed format, we may refer to its overall size as ($F \geq \acute{F}$).

In compliance with the cost model discussed in Yu et al. (2008), the energy cost of compressing source information of size F to an output of size f_o can be expressed in terms of the following function:

$$\varepsilon_{\mathrm{comp}^i_{mj(i)}}(F, f_o) = \gamma F \frac{F}{f_o}. \tag{3}$$

That is, the computation power required for compression is amplified by an increase in the compression factor (F/fo) and in proportion to the input size F. Within this cost model for data compression, the output length f_o is lower bounded by the joint entropy of all inputs contributing to the inflow of the given node together with the node's data itself, i.e., $f_o \geq H_n$. A practical notion for joint entropy modelling has been introduced in Goel and Estrin (2005).

Since the energy consumed for transmitting 1 bit is typically about 500–1000 times greater than a single 32-bit computation (Raghunathan et al., 2002), the practical meaning behind the exemplary case of $\gamma = 0.1$ is that around 50–100 instructions need to be executed for generating each bit in the output. Hence, in computation-bound scenarios, dealing with a large volume of data, simplifying assumptions for neglecting the computation overhead or taking up radical measures such as always going for maximum compression tend to be inefficient.

Continuing with our discussion of energy cost model for the compressed-communication scenario, let the node's latest view on the relevant portion of its path to the sink be characterised by its current cost and depth in the data-gathering tree. Then, $cost_i = cost_{parent(i)} + w_e$ and $depth_i = depth_{parent(i)} + 1$. Supposing that node i is delegated with mission mj, where $mj(i)$ prescribes a 'relay' role for i, it estimates the overall energy required for the preparation and transmission of compressed data of size F from an uncompressed backlog of size F along the remaining portion of the path to the sink, denoted by $\varepsilon_{\mathrm{COMP}}$ as follows:

$$\varepsilon_{\mathrm{COM}^i_{mj(i)}}(\acute{F}, F, f_o) = \sum_{\mathrm{type}(fi)=\beta} \varepsilon_{\mathrm{decomp}}(f_i) + \varepsilon_{\mathrm{comp}^i_{mj(i)}}(F, f_o) + \varepsilon_{TX^i_{mj(j)}}(f_o) + \varepsilon_{\mathrm{RCV}}(f_o) \times \mathrm{depth}_i + \varepsilon_{\mathrm{decomp}}(f_o) \tag{4}$$

In effect, the overall estimated energy accounts for the costs of: decompressing the compressed portion of the inflows, compressing the backlog, the actual transmission of data across all links on the remaining portion of the path to the sink, the receipt of data by all intermediate decision-making nodes together with the decompression overhead imposed on the immediate next node along the path.

Using equation (3) and given that $\varepsilon_{TX^i_{mj(i)}}(f_o) = f_o \times \mathrm{cost}_i$, the $\varepsilon_{\mathrm{COM}}$ equation in (4) can be rewritten as:

$$\varepsilon_{COM^i_{mj(i)}}(F, f_o) = \sum_{\mathrm{type}(fi)=\beta} \varepsilon_{\mathrm{decomp}}(f_i) + \gamma F \frac{F}{f_o} + f_o * \cos t_i + \varepsilon_{\mathrm{RCV}}(f_o) \times depth_i + \varepsilon_{decomp}(f_o) \tag{5}$$

Assuming that the reception cost for data of unit size at any given node is α, $\varepsilon_{RCV}(f_o)$ in equation (5) can be calculated as: $\alpha \times f_o$; also, given that techniques such as *gzip* consume very little time for decompression compared with compression (Barr and Asanovic, 2006; Yu et al., 2008; Tsiftes et al., 2008), we may safely abstract away the decompression cost components to render $\varepsilon_{\mathrm{COM}}$ into a convex function with a global minimum. Hence, the optimisation problem faced by node i with a decision-making mission mj can be formulated as follows:

minimize:

$$\varepsilon_{\mathrm{COM}^i_{mj(i)}}(F, f_o)$$

subject to

feasibility constraints imposed by the compression algarithm

With the first-order condition, the minimum value for the computation cost occurs at:

$$\frac{\partial \varepsilon_{\mathrm{COM}^i_{mj(i)}}(F, f_o)}{\partial f_o} = 0 \Rightarrow f_{\mathrm{opt}} = F \times \sqrt{\frac{\gamma}{\mathrm{cost}_i + \alpha \times \mathrm{depth}_i}}. \tag{6}$$

Clearly, to perform its cost analysis, a decision-making node i will compute $\varepsilon_{\mathrm{COM}^i_{mj(i)}}(F, f_o)$ for $f_o = max\ \{f_{\mathrm{opt}}, H_n\}$.

In Figure 2, we have plotted $\varepsilon_{\mathrm{COM}}$ for $\alpha = 0$, $\gamma = 0.1$, $F = 1$, and $fo \in [0.1,1]$.

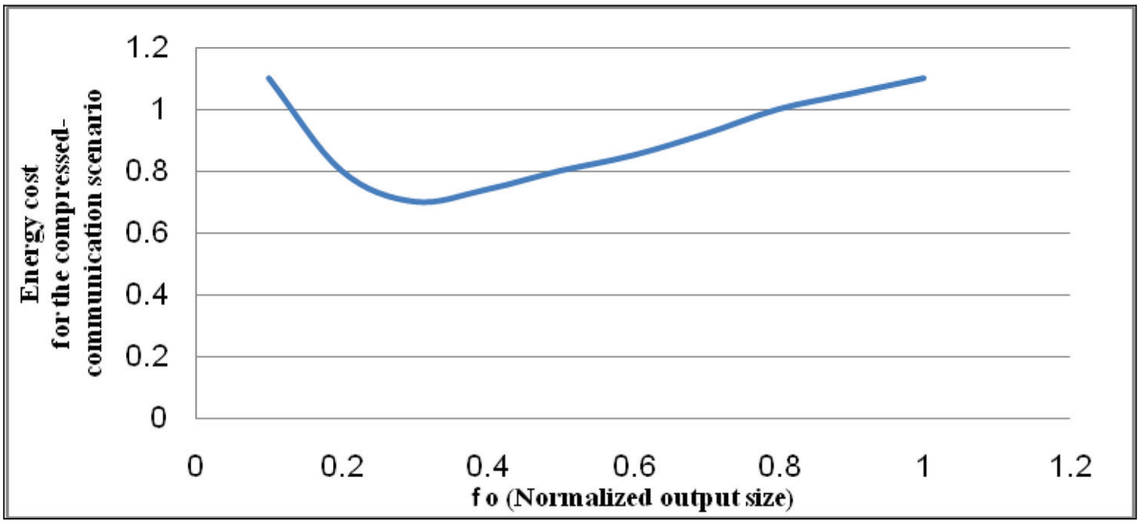
3.3 Energy cost model for the raw-communication scenario

A node i with a decision-making mission mj estimates the communication cost associated with intact relaying of the received data as follows:

$$\varepsilon_{\mathrm{RAW}^i_{mj(i)}}(\acute{F}) = \varepsilon_{TX^i_{mj(i)}}(\acute{F}) + \varepsilon_{\mathrm{RCV}}(\acute{F}) \times \mathrm{depth}_i \tag{7}$$

$$\text{where}: \varepsilon_{TX^i_{mj(i)}}(F') = F' * \mathrm{cost}_i \text{ and } \varepsilon_{\mathrm{RCE}}(\acute{F}) = \alpha.$$

Figure 2 Energy cost for the compressed communication scenario (see online version for colour)



3.4 The decision-making process

In each data aggregation round, every relay node delegated with a decision-making mission is supposed to weigh the options of transmitting with or without compression with respect to the incurred end-to-end costs. In effect, each node endeavours to adaptively reduce the end-to-end energy consumption by deriving a trade-off between relaying data in compressed form using a constrained optimal compression ratio, on the one hand, and sending in the raw, on the other.

We define an indicative decision variable as follows:

$$\Delta^{i}_{mj(i)} = \varepsilon_{\mathrm{RAW}^{i}_{mj(i)}}(\acute{F}) - \varepsilon_{\mathrm{COM}^{i}_{mj(i)}}(F, f_o = \max\{f_{\mathrm{opt}}, H_n\}), \quad (8)$$

which can be written in simple form as:

$$\Delta^{i}_{mj(i)} = (F' - f_o) * (\mathrm{cost}_i + \alpha \times \mathrm{depth}_i) - \gamma \frac{F^2}{f_o}. \quad (9)$$

Hence, the decision-making process in node i with mission j comes down to drawing the following comparison:

$$\begin{cases} \text{send data in the raw,} & \Delta^{i}_{mj(i)} < 0 \\ \text{compress before sending} & \Delta^{i}_{mj(i)} \geq 0 \end{cases}. \quad (10)$$

If we set $f_o = f_{opt}$ (from equation (6)),

$$\varepsilon_{\mathrm{RAW}^{i}_{mj(i)}}(\acute{F}) \geq \varepsilon_{\mathrm{COM}^{i}_{mj(i)}}(F, f_{\mathrm{opt}}) \Leftrightarrow$$

$$F' \times (\mathrm{cost}_i + \alpha \times \mathrm{depth}_i) \geq \gamma \frac{F^2}{f_{\mathrm{opt}}}$$

$$f_{\mathrm{opt}} + (\mathrm{cost}_i + \alpha \times \mathrm{depth}_i)$$

$$\xleftrightarrow{\;f_{\mathrm{opt}} = F\sqrt{\frac{\gamma}{\mathrm{cost}_i + \alpha \times \mathrm{depth}_i}}\;} F' \geq 2 f_{\mathrm{opt}}.$$

Finally, it can easily be seen that having every decision-making node on a given data-gathering path strike a trade-off between the communication and computation costs, the actual energy consumption for delivering the load associated with a node i_1 at the extreme downstream end of the path is guaranteed to be upper bounded by: $\min\left(\varepsilon_{\mathrm{RAW}^{i_1}_{mj(i_1)}}, \varepsilon_{\mathrm{COM}^{i_1}_{mj(i_1)}}\right)$.

4 Performance evaluation

We have conducted packet-level simulation experiments using OMNET++ (Varga, 2001) to obtain preliminary performance measurement results. In this section, we report on the evaluation of our proposal, AREM, against a comparable scheme from prior art, the so-called TCDG algorithm (Yu et al., 2008). For the sake of the experiments, it has been assumed that the sensor nodes are deployed uniformly across a 3200×3200 two-dimensional field that the sink node is placed at the bottom-left corner of the field, and that each link is weighted as cd^2, where d denotes the distance between two incident nodes and $c = 2 \times 10^{-5}$ J/m^2. AREM and TCDG have been evaluated over both Shortest Path Tree (SPT) as well as Minimal Steiner Tree-based (MST) data-gathering substrates, and each data point is calculated by averaging over 40 runs reported within 95% confidence interval. As for the joint entropy models used throughout the experiments, we assume a stationary Gaussian random process with a scalar quantiser, uniform step size, and infinite number of levels (Marco et al., 2003). Given that our envisioned set-up in this paper mimics the case with TCDG, we encourage the interested reader to refer to Yu et al. (2008) for more specific details.

4.1 Impact of the data entropy rate (ρ)

For the purpose of the first scenario, we consider a varying data entropy rate within the range [0, 0.4], and will investigate how the WSN's energy usage behaviour will change with respect to the expected value of the information contained in messages. Figure 3 demonstrates the outcome of this experiment for $N = 200$, $\gamma = 0.1$, and with 50% of the nodes acting as sources (i.e., $|R| = 100$).

As can be seem in Figure 3, for larger values of ρ, there is a continuously increasing requirement for joint compression of data and owing to its lower cost of data transmission, MST exhibits a better performance; contrarily, for small values of ρ, SPT supersedes. In general, with ρ increasing, the information volume in the network grows and the overall energy consumption also increases to handle the transmission of the data. The optimal value for ρ turns out to be 0.1. Also, it is noticeable that AREM achieves a lower level of power consumption compared with TCDG thanks to the inclusion of the upstream inflows into its decision-making mechanism.

4.2 Impact of the relative computation cost (γ)

To investigate the impact of the relative computation cost on the network's energy usage profile, we vary γ within the range [0, 0.24] with $N = 200$, $\rho = 0.1$, and $|R| = 100$. As can be noticed in Figure 4, for these low computation costs, the discrepancy between the two types of trees is not very remarkable for both algorithms; however, the relative superiority of SPT over MST can be attributed to the fact that for larger computation costs, the compression ratio reduces, thus rendering the shortest paths to the sink more preferable.

In general, with the relative computation cost (γ increasing, more energy is consumed in the network. As expected, by reducing the number of source nodes, the difference in costs incurred by each algorithm with respect to its underlying tree structure would become less pronounced (see Figure 5 for $|R| = 50$).

4.3 Impact of the number of source nodes ($|R|$)

In Figure 6, the percentage of source nodes is varied to study the impact of the number of data generators on energy consumption. It has been assumed that $\gamma = \rho = 0.1$.

Figure 3 The impact of the data entropy rate (see online version for colour)

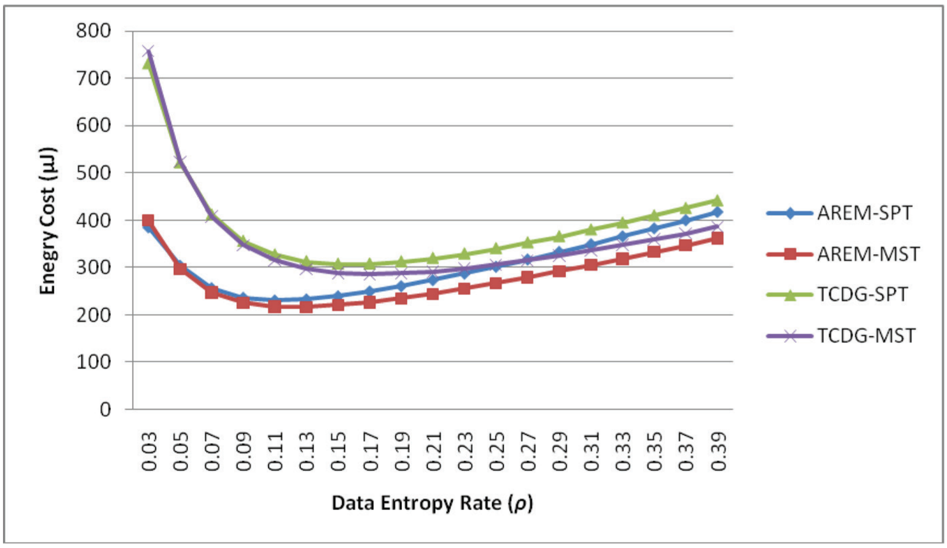


Figure 4 The impact of the relative computation cost (|R| = 100) (see online version for colour)

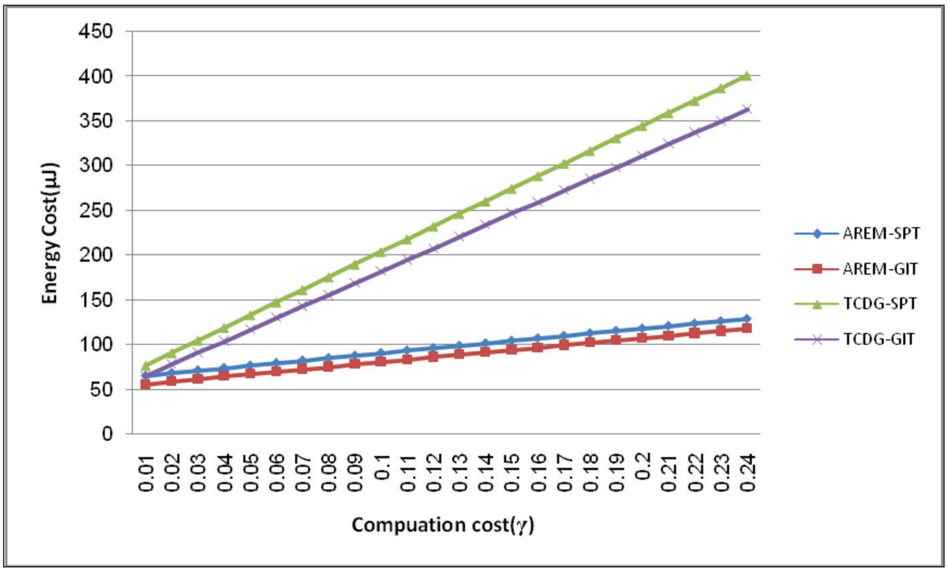


Figure 5 The impact of the relative computation cost (|R| = 50) (see online version for colours)

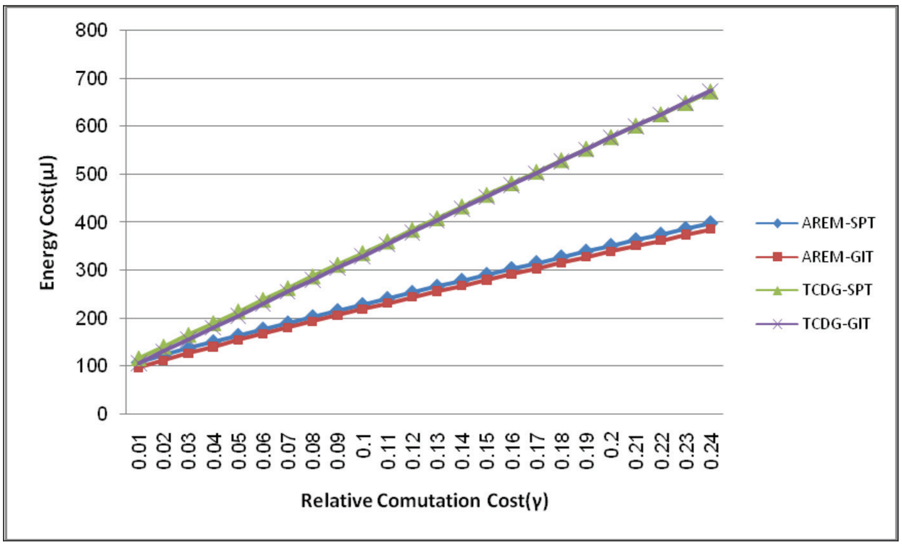


Figure 6 The impact of the percentage of the source nodes (see online version for colours)

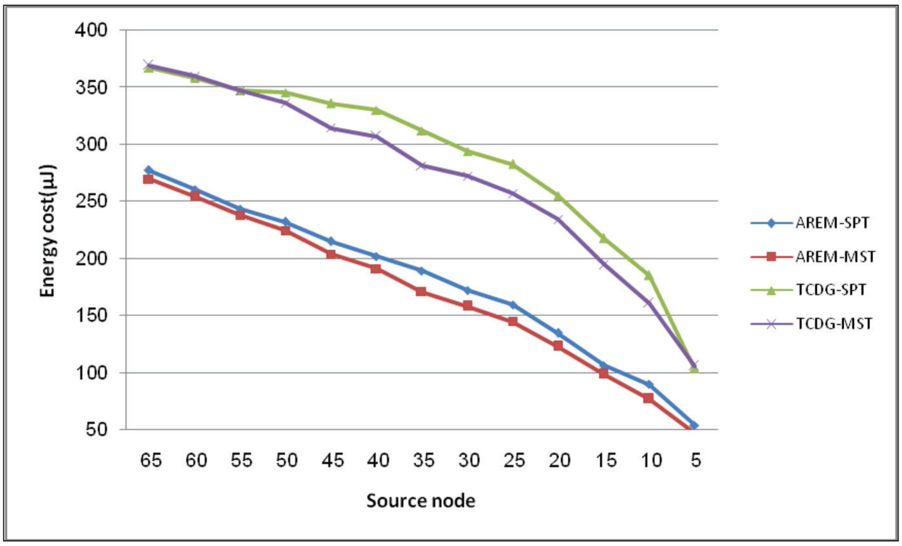
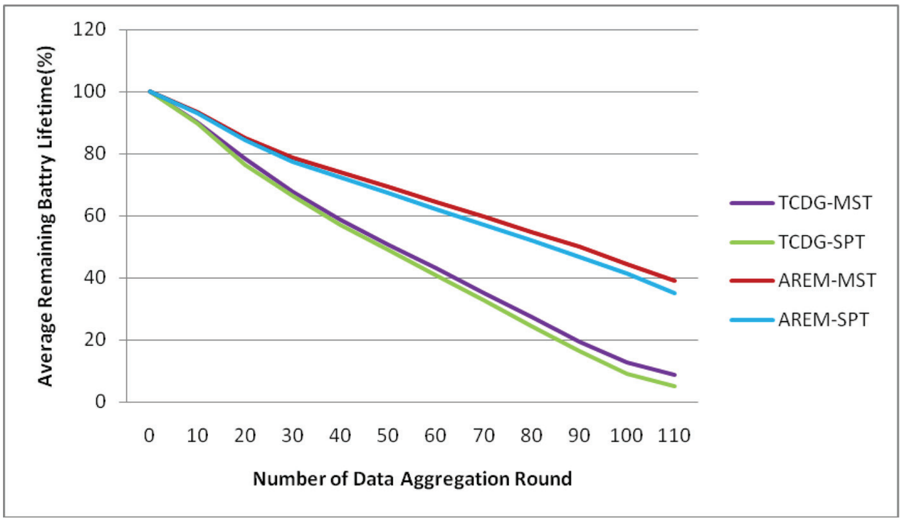


Figure 7 Network lifetime in terms of the number of data aggregation rounds (see online version for colours)



The network's energy depletes almost linearly as the number of source nodes increases. However, the depletion scale is smaller in our scheme compared with the case with TCDG given that AREM takes the initiative to regulate the actions of downstream nodes so that a minimum upper bound is ensured for the end-to-end energy costs.

4.4 Network lifetime in terms of the number of the data aggregation rounds

We have measured the number of rounds it takes for both algorithms to use up the network's energy by setting $N = 200$, $\rho = \gamma = 0.1$, and with 50% of the nodes acting as sources (i.e., $|R| = 100$).

As can be seen in Figure 7, AREM running on MST turns out to present with the most favourable result. Also of note in the figure is the marginal superiority of MST over SPT as the underlying data-gathering structure in both algorithms. Once again, the slower depletion rate in our scheme is attributable to the fact that AREM is not oblivious to the upstream inflows and unlike TCDG, the cost analysis at a node is conducted with respect to the impact it might have on the decisions of the succeeding nodes along the path.

5 Conclusions

Recent trends in energy-aware techniques for WSNs have presented a caveat to the extremist application of compression algorithms for cutting down on communication overhead. It has been demonstrated that neither maximum compression nor avoiding data compression completely can lead to utmost network longevity; instead, the ideal policy lies in between these two pure strategies and should be adjusted dynamically. In this paper, we have also examined the ramifications of lossless compression of data prior transmission over a data-gathering tree in terms of the resultant energy costs. We adaptively reduce the end-to-end energy consumption by deriving an online trade-off between relaying data in compressed form with a constrained optimal compression ratio, on the one hand, and sending in the raw, on the other. Our work differs from the previous research in that the notion of tunable compression adopted in this paper is not oblivious to the upstream inflows and that the aggregation process is governed dynamically with the backlog of all contributing sources including the current node's as well. We also proffer a more accurate model of compression cost when compared with the prior art and relax the assumption that limits the sources to the leaves of the data-gathering tree. Simulation results are provided to quantify the performance of our proposed scheme. In most cases, the outcome of the experiments reveals promising performance by almost 36% additional reduction in energy costs.

References

Acimovic, J. Beferull-Lozano, B. and Cristescu, R. (2005) 'Adaptive Distributed Algorithms for Power-Efficient Data Gathering in Sensor Networks,' *in Symp. Wireless Sensor Networks*, Vol. 2, June 2005, pp.946–951.

AL-khdour, T. and Baroudi, U. (2009) 'A generalised energy-efficient time-based communication protocol for Wireless Sensor Networks', *International Journal of Internet Protocol Technology*, Vol. 4, No. 2, pp.134–146.

Barr, K. and Asanovi, K. (2003) 'Energy aware lossless data compression', in *Mobile Systems, Applications, and Services (MobiSys'03)*, Vol. 24, San Francisco, CA, USA, May, pp.231–244.

Barr, K.C. and Asanovic, K. (2006) 'Energy-aware lossless data compression', *ACM Transactions on Computer Systems*, Vol. 24, No. 3, pp.250–291.

Choi, W. and Das, S.K. (2005) 'A novel framework for energy-conserving data gathering in wireless sensor networks', in *The 24th Conference of the IEEE Communications Society (IEEE INFOCOM 2005)*, Vol. 3, Miami, Florida, USA, March, pp.1395–1404.

Dong, W., Liu, X., Chen, C., He, Y., Chen, G., Liu, Y. and Bu, J. (2010) 'DPLC: dynamic packet length control in wireless sensor networks', *INFOCOM 2010, San Diego, CA, ISSN 0743-166X*, March, pp.1–9.

Eswaran, S., Johnson, M., Misra, A. and Porta, T. (2009) 'Adaptive in-network processing for bandwidth and energy constrained mission-oriented multi-hop wireless networks', *Proc. of DCOSS*, Los Angeles, CA, June.

Fasolo, E.M., Rossi, J., Widmer and Zorzi, M. (2007) 'In-network aggregation techniques for wireless sensor networks: a survey', *IEEE Wireless Communications*, Vol. 14, No. 2, pp.70–87.

Goel, A. and Estrin, D. (2005) 'Simultaneous optimization for concave costs: single sink aggregation or single source buy-at-bulk', *in Symp. Discrete Algorithms (SODA '03)*, Vol. 43, pp.5–15.

Haque, M.E., Matsumoto, N. and Yoshida, N. (2009) 'Context-aware cluster-based hierarchical protocol for Wireless Sensor Networks', *International Journal of Ad Hoc and Ubiquitous Computing*, Vol. 4, No. 6, pp.379–386.

Hong, B. and Prasanna, V.K. (2004) 'Optimizing a class of in-network processing applications in networked sensor systems', in *Mobile Ad Hoc and Sensor Systems (MASS '04)*, October, pp.154–163.

Kanna, R. and Iyengar, S.S. (2004) 'Game-Theoretic models for reliable path-length and energy-constrained routing with data aggregation in wireless sensor networks', in *IEEE J. Select. Areas Commun*, Vol. 22, August, pp.1141–1150.

Leone, P., Nikoletseas, S.E. and Rolim, J. (2005) 'An adaptive blind algorithm for energy balanced data propagation in wireless sensors networks', *in Distributed Computing in Sensor Systems (DCOSS)*, Vol. 3560, Berlin, June, pp.35–48.

Liang, W., Member, S. and Liu, Y. (2007) 'Online data gathering for maximizing network lifetime in sensor networks', *IEEE Transactions on Mobile Computing*, Vol. 6, January, pp.2–11.

Marco, D., Duarte-Melo, E.J., Liu, M. and Neuhoff, D.L. (2003) 'On the many-to-one transport capacity of a dense wireless sensor network and the compressibility of its data', in *Information Processing in Sensor Networks (IPSN '03)*, April, pp.1–16.

Naderan, M., Dehghan, M., Pedram, H. and Hakami, V. (2012) 'Survey of mobile object tracking protocols in Wireless Sensor Networks: a network-centric perspective', *International Journal of Ad Hoc and Ubiquitous Computing*, In Press.

Raghunathan, V. Schurgers, Park, C.S. and Srivastava, M.B. (2002) 'Energy-aware wireless microsensor networks', *IEEE Signal Processing Magazine*, Vol. 19, May, pp.40–50.

Tahir, M. and Farrell, R. (2009) 'Optimal communication-computation tradeoff for wireless multimedia sensor network lifetime maximization', *Wireless Communications and Networking Conference*, WCNC 2009, ISSN 1525-3511, Budapest, May, pp.2514–2519.

Tsiftes, N., Dunkels, A. and Voigt, T. (2008) 'Efficient sensor network reprogramming through compression of executable modules', in *The 5th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON'08)*, pp.359–367.

Varga, A. (2001) 'The OMNeT++ discrete event simulation system', in *Proceedings of the European Simulation Multiconference (ESM'01)*.

Yu, Y., Krishnamachari, B. and Prasanna, V.K. (2008) 'Data gathering with tunable compression in sensor networks', *IEEE Transactions on Parallel and Distributed Systems*, Vol. 19, February, pp.276–287.

Zhu, Y. Sundaresan, K. and Sivakumar, R. (2005) 'Practical limits on achievable energy improvements and useable delay tolerance in correlation aware data gathering in wireless sensor networks', *in Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON)*, Santa Clara, CA, US, September.

Zytoune, O., Aroussi, M.E. and Aboutajdine, D. (2011) 'An energy efficient clustering protocol for routing in Wireless Sensor Network', *International Journal of Ad Hoc and Ubiquitous Computing*, Vol. 7, No. 1, pp.54–59.