

# Data Aware Communication for Energy Harvesting Sensor Networks

Mohamed Hefeida, Fahad Saeed

► **To cite this version:**

Mohamed Hefeida, Fahad Saeed. Data Aware Communication for Energy Harvesting Sensor Networks. 14th International Conference on Wired/Wireless Internet Communication (WWIC), May 2016, Thessaloniki, Greece. pp.121-132, 10.1007/978-3-319-33936-8\_10 . hal-01434847

**HAL Id: hal-01434847**

**<https://hal.inria.fr/hal-01434847>**

Submitted on 13 Jan 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Data Aware Communication for Energy Harvesting Sensor Networks

Mohamed Hefeida<sup>1</sup> and Fahad Saeed<sup>2</sup>

<sup>1</sup> American University of the Middle East  
AlEqaila, Kuwait

`mohamed.hefeida@aum.edu.kw`

<sup>2</sup> Western Michigan University  
Kalamazoo, Michigan 49008  
`fahad.saeed@wmich.edu`

**Abstract.** We propose a Data Aware Communication Technique (DACT) that reduces energy consumption in Energy Harvesting Wireless Sensor Networks (EH-WSN). DACT takes advantage of the data correlation present in household EH-WSN applications to reduce communication overhead. It adapts its functionality according to correlations in data communicated over the EH-WSN and operates independently from spatial and temporal correlations without requiring location information. Our results show that DACT improves communication efficiency of sensor nodes and can help reduce idle energy consumption in an average-size home by up to 90% as compared to spatial/temporal correlation-based communication techniques.

**Keywords:** Sensor networks, energy harvesting, energy efficiency, data collection, data redundancy

## 1 Introduction

With continuous and rapid advancements in microelectronics and wireless communications, smart devices have become an integral and critical part of our everyday lives. From smart phones to smart fridges, almost every part of our daily routine involves an electrical/electronic device that consumes some form of energy, mainly electricity. However, due to various behavioral, technological, and social reasons, most of us do not have a clear idea of the amount of energy required to sustain our daily habits. Part of this is due to the lack of a clear and real-time measure for the effect of our daily habits and behavior on energy usage. An example of this lack of awareness is leakage power (a.k.a. standby power), which results from leaving electrical devices connected to the power outlet. According to [13], the average US household loses \$100 every year on leakage power. On a national level these losses reach up to US\$ 100 billion and a carbon footprint equivalent to 26.2 million tons of CO<sub>2</sub> emissions in the US alone.

Studies have shown that about 71% of consumers are willing to change their energy-related habits if they had clear information about their real-time energy

usage/cost [13]. Therefore, a quantitative method is required to raise awareness among consumers about their energy consumption habits.

The continual scaling of transistors allows today's electronic circuits to operate at a fraction of the power of their counterparts a few years ago, allowing electronics to be completely powered by ambient energy. In addition, wireless communication technology has grown exponentially, allowing unprecedented wireless communication amongst devices & access to the Internet through various protocols (e.g. Wifi, ZigBee, Bluetooth). The combination of these two main advancements in microelectronics and wireless communication paves the road for developing Energy Harvesting Wireless Sensor Networks (EH-WSN) that partially or completely run on scavenged energy [13]. An EH-WSN can harvest power from surrounding environments while consuming very little power. Therefore, optimization of power consumption at the lowest possible level is essential for EH-WSN to operate by using idle power from other devices.

Like other WSN applications, an EH-WSN involves data gathering, in-network information processing and data aggregation (e.g. [4, 9, 10, 21]). However, the EH-WSN application at hand has a unique feature, where correlations in the data space are not necessarily due to spatial and/or temporal correlations in the sensed phenomenon. Therefore, communication schemes that take advantage of spatial and/or temporal correlations [17], [12] will not be effective in the energy monitoring EH-WSN application at hand. Physical phenomena usually result in similar data due to spatially and/or temporally correlated fields (e.g. monitoring temperature), but this may not necessarily be true for energy usage monitoring. For example, a household where one power outlet is idle while the one right next to it is plugged to a power-hungry microwave is an example of spatial correlation that does not lead to data correlation. We take advantage of this independence in data, temporal, and spatial correlations in EH-WSN and propose a communication scheme that will selectively communicate data based on its significance compared to other data.

In this paper, we tackle the problem of redundant data communication in EH-WSN from a collaborative communication perspective and evaluate the operation of the Information Processing and Communication Reduction (IPCR) scheme proposed in [10]. In contrast to other schemes [6, 7, 9, 12, 17], [10] relies only on data similarity rather than on underlying field correlations. It employs a clever mechanism that compares a node's current sensed data to that communicated over the channel, based on which a decision of transmission or suppression of data is made. That is, if each node processes the information transmitted over the channel (by other nodes) to check its similarity with other sensed data, it can make a more informed decision on whether to transmit its sensed value.

The remainder of this paper is organized as follows: Section 2 is an overview of related studies that focus on reducing energy consumption in WSN. In 3, we map the problem to a well known collaborative sequential spectrum sensing problem based on which our problem is formulated. Section 4 describes the proposed scheme and the effect of data similarity on its operation. Finally, we evaluate the proposed scheme in section 5, followed by the conclusion in section 6.

## 2 Related Work

The majority of research efforts in improving energy efficiency of EH-WSN focus on communication operations. This is due to that communication operations consume orders of magnitude more energy as compared to computation operations (e.g. 2000X [14]). However, information processing can greatly improve the significance of the communicated data.

Few investigations have focused on joint solutions to MAC schemes and information processing [17], [12]. Our previous work in [10] proposed a joint solution to reduce communication operations via collaborations between the underlying MAC scheme and a field estimation technique. We take advantage of [10] and modify it to be utilized in our EH-WSN application.

## 3 Problem Statement and Assumptions

### 3.1 Overview

Consider a household EH-WSN where standby power is monitored by a group of electro-magnetic radiation sensor nodes. Each node is required to report its sensed idle power to a central node (sink). The sink is connected to the Internet and is responsible for processing the results of all nodes and reporting energy consumption data sending alerts to a smart phone when a preset threshold of consumption is detected. Assume a total of  $N_p$  electro-magnetic radiation sensor nodes monitoring power outlets in the EH-WSN, operating in a non-data aware manner (i.e. each node is unaware of other data but its own), each of the  $N_p$  nodes transmits its sensed data independently to the sink node. According to the *similarity* of the sensed field, this can result in up to  $N_p - 1$  redundant (*similar*) messages. Moreover, many collaborative communication reduction techniques tie their performance to underlying field correlations [6,12,17]. While this can be effective in applications encountering some spatially and/or temporally correlated fields (e.g. monitoring physical phenomena), it falls short when the underlying field encounters spatially and/or temporally uncorrelated *similar* values such as the case of a household where one power outlet is idle while the one right next to it is plugged to a power-hungry microwave.

Several studies proposing efficient collaborative communication techniques have been presented in the literature [5,22]. The main goal of these studies is to detect when the channel is not being used by primary users so that secondary users can utilize it during that time. In our problem, nodes sense the channel in order to detect ongoing transmissions and determine whether they have useful information to send. Moreover, all nodes have an equal opportunity and capability of acquiring the channel. Nodes listening to the channel during an ongoing transmission determine the relevance of their sensed values to those being transmitted over the channel. Therefore, the channel is monitored by non-transmitting awake nodes for other transmissions, based on which they decide

whether to send their data. We tailor common representations of the collaborative communication problem in [5, 10, 22] to formulate our problem and the proposed solution.

### 3.2 Formulation

Consider  $N_p$  electro-magnetic radiation sensor nodes placed at the power outlets of a home to monitor its electricity usage, within each others' communication range and reporting to a sink node. The nodes are attempting to collectively solve a binary hypothesis testing problem, where each of the  $N_p$  nodes is required to decide between transmitting its local sensed data (hypothesis  $H_1$ ) to the sink node or not (hypothesis  $H_0$ ). We assume that time is divided into discrete slots of equal durations,  $\tau_s$ , in which a node can transmit/receive data. We assume that all nodes listen to each others' transmissions (i.e. each transmission is a broadcast). The term slot and observation interval are used interchangeably. Let  $O_i(t)$  be the observed value at node  $i$  during slot  $t$ , and  $S_i(t)$  be the sensed value at node  $i$  during the same slot, where  $i = 1, 2, \dots, N_p$ . Notice that the observed value at node  $i$  is that transmitted by any of the other  $N_p - 1$  nodes in the network, while the sensed value is that sensed by node  $i$  itself. Elements constructing the sets of observed and sensed values at node  $i$  over a time span of  $T$  slots,  $\{O_i(t)\}_1^T$  and  $\{S_i(t)\}_1^T$ , respectively, are independent given each hypothesis and are assumed to be identically distributed. Equation 1 represents  $O_i(t)$  under the two hypotheses.

$$\begin{aligned} H_0 : S_i(t) - Thr_i(t) + W_i(t) &\leq O_i(t) \leq S_i(t) + \\ &Thr_i(t) + W_i(t), \quad t = 1, 2, \dots, T \end{aligned} \quad (1)$$

$H_1 : otherwise$

where  $Thr_i$  is the permissible threshold between the observed and sensed values, which reflects the level of energy consumption reporting accuracy required by the consumer.  $W_i$  is additive white Gaussian noise with a power of  $\sigma^2$ , assumed to be similar at all nodes.

As in [5, 10, 11] and without loss of generality, the primary signal  $S_i(t)$  is assumed to be a real zero-mean Gaussian random variable. Moreover, the conditional probability distributions of  $O_i(t)$  given  $H_1$  and  $H_0$  are represented by  $f_{O_i(t)}(o_t|H_1)$  and  $f_{O_i(t)}(o_t|H_0)$ , respectively. Each of them can be represented as follows:

$$\begin{aligned} f_{O_i(t)}(o_t|H_0) &\sim \mathcal{N}(0, \sigma^2) \\ f_{O_i(t)}(o_t|H_1) &\sim \mathcal{N}(0, \sigma^2 + \sigma_{s_i}^2) \end{aligned} \quad (2)$$

where  $\mathcal{N}(0, \sigma^2)$  and  $\mathcal{N}(0, \sigma^2 + \sigma_{s_i}^2)$  are normal distributions with zero means and variances of  $\sigma^2$  and  $\sigma^2 + \sigma_{s_i}^2$ , respectively.  $\sigma_{s_i}^2$  represents the average received primary signal at the  $i^{th}$  node, which is assumed to be fixed over the time slot duration [11]. Therefore, any local observation at a node  $i$  can be expressed as:

$$Y_i = \sum_{t=1}^T \log \left[ \frac{f_{O_i(t)}(o_t|H_1)}{f_{O_i(t)}(o_t|H_0)} \right] \quad (3)$$

where  $Y_i$  is the Log-Likelihood Ratio (LLR), computed by node  $i$  [22]. The signal-to-noise ratio (SNR) can be defined as  $\psi_i = \sigma_{o_i}^2 / \sigma^2$ . This will result in an LLR computed at node  $i$  as follows:

$$Y_i = \frac{\psi_i}{2\sigma^2 + 2\psi_i\sigma^2} \sum_{t=1}^T |O_i(t)|^2 - \log(1 + \psi_i) \frac{T}{2} \quad (4)$$

Both [5, 11] propose approximations for the above likelihood functions of  $Y_i$  given either  $H_0$  or  $H_1$ , which are shifted scaled chi-square distributions with  $T$  degrees of freedom.

## 4 Proposed Technique

In this section, we describe the details of DACT, its realization of the Information Processing and Communication Reduction (IPCR) scheme presented in [10] and study the effect of data similarity on its operation.

### 4.1 Operation

Consider the same EH-WSN in section 3. Each node  $i \in N_p$ , sets a backoff (BO) timer according to its locally computed LLR, such that  $BO \propto 1/|LLR|$ . Moreover, each node  $i$  checks the condition given in (1) to determine whether it will transmit or not. That is, if a node determines that it has highly informative information, based on the value of the BO timer (reflecting the LLR value), but it doesn't satisfy the condition in (1), it will decide *not* to transmit. This will repeat  $\forall i \in N_p$ . Table 1 shows the pseudo code of DACT's utilization of IPCR, the details of IPCR operation have been omitted and can be found in [10].

Each node senses the field and sends its value if the channel is sensed idle (i.e. empty). If the channel is sensed busy, a node compares its sensed value and that being communicated over the channel according to (1). Note that this a comparison of the received data (communicated over the channel) and that locally sensed by the node. Based on this comparison, a node either decides to send ( $H_1$ ) or discard ( $H_0$ ) its data. The degree of information accuracy is controlled via the threshold ( $Thr_i$ ) set by the user according to their habits/preference. That is, if users set the threshold to a lower value, that will result in more communication and hence less energy savings and vice versa. A detailed example of DACT's utilization of IPCR is discussed in section 5.

To avoid deadlocks, if none of the  $N_p$  nodes in the network transmits in communication round  $r$ , the first node to acquire the channel in communication round  $r + 1$  will transmit its locally sensed value regardless of the current similarity check as in (1).

### 4.2 Communication Cost

The operation of DACT requires evaluation of neighbors' transmitted data, which involves significant communication overhead. We identify different sources

**Table 1.** Basic DACT-IPCR Operation

---



---

**Initialization:**  $\forall i \in N_p$

1.  $iter \leftarrow 0$  /\*set counter\*/
2.  $Thr_i \leftarrow Threshold$  /\*set Threshold\*/

**Begin**

3.  $S_i \leftarrow sensed\ value$
4. *listen to channel*
5. *If channel is idle*
6.     *transmit  $S_i$*
7. *Else*
8.      $O_i \leftarrow ongoing\ transmission$
9.     *If ( $|O_i - S_i| \leq Thr_i$ ) /\*according to (1)\*/*
10.     *discard  $S_i$*
11.     *exit*
12. *Else*
13.     Repeat
14.          $iter++$
15.         *goto line 4*
16.     Until  $iter \leq T$  /\* $T$  is the max number of slots in any frame\*/

**End**

---

of energy consumption and their dependence on network and application parameters, such as collisions and queue utilization. Our analysis is based on common channel assumptions that have been used in the literature [6, 9, 10].

In duty-cycled MAC schemes for WSNs, there are two main states for a node: active and sleep. During its active state, a node can transmit, receive or listen to the channel. The node turns off its radio during its sleep state. Each node is assumed to have a queue of finite length  $Q$ , and each packet in the queue has an average length  $L_{DATA}$  bits. We assume that data packets are generated following a Poisson process with a rate equal to  $\lambda$  packets/second (i.e. inter-packet times are independent and have an exponential distribution with a mean  $= 1/\lambda$ ). However, more complex traffic models can also benefit from our technique but with different distributions of trade-offs between energy consumed in channel sensing and that saved from collision avoidance and conditional message transmissions. Each packet is assumed to spend an average of  $T_{delay}$  before leaving the queue, which is the sum of queuing delay and service time computed by:

$$T_{delay} = \tau_s N_s + (A - 1)(\tau_s N_s + T_C) \quad (5)$$

where  $\tau_s$  is the average slot duration,  $N_s$  is the average number of slots skipped before acquiring the channel on each transmission attempt (Back off window),  $A$  is the average number of transmission attempts needed per packet, and  $T_C$  is a collision duration.  $A$  can be represented as a function of the collision probability  $P_C$  and the maximum number of retransmission attempts  $R_A$  such that  $A =$

$\frac{(1-P_C^{(R_A+1)})}{1-P_c}$ ). The collision probability is related to the number of nodes in the network,  $N_p$ , where  $P_C = 1 - (1 - P_r)^{N_p-1}$  and  $P_r$  is the probability of a node, having a packet ready to be sent, to transmit in a random slot.  $P_r$  can be related to the queue utilization factor  $\rho$  by  $P_r = \rho/(N_s + 1)$ , where  $\rho = \lambda/\mu$  and  $\mu$  is the mean service time.  $T_C$  is deduced from IEEE 802.11 as well as the values of the guard periods, SIFS (Short Inter-Frame Space), DIFS (Distributed Inter-Frame Space), and EIFS (Extended Inter-Frame Space) [18].  $T_C = DIFS + SIFS + L_{RTS}/r$  and  $T_S = \frac{L_{RTS}+L_{CTS}+L_{DATA}+L_{ACK}}{r} + DIFS + 3SIFS$  is the time needed to successfully transmit one data packet. Note that for a uniformly distributed back off window over the maximum contention window will lead to  $N_s = \frac{CW_{max}}{2}$ . The throughput of the queue can be computed as  $\gamma = \lambda(1 - P_B)$  and  $P_B = \frac{(1-\rho)\rho^Q}{1-\rho^{Q+1}}$  is the blocking probability (i.e. probability that the buffer is full).

We assume possible channel states with respect to the sending node to be: (a) empty (neighbor nodes are idle listening or sleeping), (b) sending/receiving, and (c) collision. Each state has a corresponding probability of (a)  $P_e = (1 - P_r)^{N_p-1}$ , (b)  $P_{s/r} = P_r(N_p - 1)(1 - P_r)^{N_p-2}$ , and (c)  $P_c = 1 - P_{s/r} - P_e$ , respectively. The total energy consumption of a node is due to transmitting, receiving, and overhearing. Each one of these energy components has a certain successful and collision component in it. This leads to:

$$E_{total} = E_{tx}^s + E_{tx}^c + E_{rx}^s + E_{rx}^c + E_{oh}^s + E_{oh}^c \quad (6)$$

where  $E_{tx}^s$ ,  $E_{rx}^s$ , and  $E_{oh}^s$  are the energies consumed in successful transmission, reception and overhearing, respectively.  $E_{tx}^c$ ,  $E_{rx}^c$  and  $E_{oh}^c$  are the energies consumed in collided (unsuccessful) transmission, reception and overhearing, respectively. The value of each one of these energy components will vary according to the MAC protocol behavior. Each of the above energy components suffers an amount of idle listening as well (e.g., during DIFS and SIFS). We refer to the energy consumed in a node's radio states, transmission, reception and idle as  $E_{radio}^{TX}$ ,  $E_{radio}^{RX}$ , and  $E_{radio}^{IDLE}$ , respectively. Radio sleep state is assumed to consume no energy, therefore:

$$E_{tx}^s = E_{radio}^{TX} \frac{L_{RTS} + L_{DATA}}{r} + E_{radio}^{RX} \frac{L_{CTS} + L_{ACK}}{r} + E_{radio}^{IDLE} (DIFS + 3SIFS + N_s P_e \epsilon) \quad (7)$$

$$E_{tx}^c = E_{radio}^{TX} \frac{L_{RTS}}{r} + E_{radio}^{IDLE} (DIFS + 2SIFS + N_s P_e \epsilon + \frac{L_{CTS}}{r}) \quad (8)$$

where  $\epsilon$  is the duration of an empty slot. The energy consumed in successful and unsuccessful receptions can be represented by (9) and (10), respectively.



$$E_{rx}^s = E_{radio}^{RX} \frac{L_{RTS} + L_{DATA}}{r} + E_{radio}^{TX} \frac{L_{CTS} + L_{ACK}}{r} + E_{radio}^{IDLE} (3SIFS) \quad (9)$$

$$E_{rx}^c = E_{tx}^c - (E_{radio}^{TX} - E_{radio}^{RX}) \frac{L_{RTS}}{R} \quad (10)$$

### 4.3 Data Similarity

In order to offset the communication overhead encountered in DACT, a certain level of data similarity is required. Fortunately, this is the usual case when people are not at home and standby power is being wasted. DACT utilizes data similarity to detect and reduce redundant information communicated over the channel. We have used a similar scheme to exploit data redundancy in genomic data for efficient transmission [2]. Note that this distinguishes the energy consumption monitoring EH-WSN applications from other WSN applications, since similarity in sensed values does not necessarily reflect any spatial correlation, as discussed in section 3.1. To represent data similarity, like in previous studies [9, 10], we define a similarity factor  $F_p = \frac{N_p}{N_S}$ ,  $1 \leq N_S \leq N_p$ , where  $N_S$  is the number of sets representing the field which has a total of  $N_p$  nodes and  $S_{V_x}$  is the set of nodes in a neighborhood with a sensed field value  $V_x$ ,  $min(f) \leq V_x \leq max(f)$ , where  $min(f)$  and  $max(f)$  are the minimum and maximum values of the sensed field, respectively. Note that  $\forall V_x \neq V_y, S_{V_x} \cap S_{V_y} = \emptyset$ .

### 4.4 Discussion

Consider a building with 100 power outlets that are being monitored for idle power consumption via electro-magnetic radiation sensor nodes and a sink node, forming an EH-WSN. Assume that all sensors are calibrated to sense power on a scale of 1 to 100. At one extreme, all 100 nodes sense 90 and thus belong to one set  $S_{90}$ , therefore  $F_p = N_p$ , which is the maximum value for  $S$  in a neighborhood of size  $N_p$ . At the other extreme, all 100 nodes report 100 different values which results in 100 different sets and thus  $F_p = 100/100 = 1$ , which is the minimum possible value for  $F_p$  indicating no similarity in the sensed field.

DACT's strength lies in exploiting data similarity and redundancy that is absent in other studies [6, 12, 17]. In a best case scenario, when the field is highly similar (highly correlated in data space and thus low values of  $F_p$ ), significant communication savings are expected, as illustrated by the previous example. However, if the field is highly dissimilar (high values of  $F_p$ ), DACT will perform comparable to field correlation-based collaborative scheme [6, 12, 17] or non-collaborative data gathering/aggregation techniques [20]. We refer to both as conventional techniques.

**Table 2.** Simulation Parameters

Parameter	Value	Parameter	Value
<i>Bandwidth</i>	20 kbps	<i>Comm. Range</i>	250 m
<i>RxPower</i>	22.2 mW	<i>Interference Range</i>	550 m
<i>TxPower</i>	31.2 mW	<i>DIFS</i>	10 ms
<i>IdlePower</i>	22.2 mW	<i>SIFS</i>	5 ms
<i>SleepPower</i>	3 $\mu$ W	<i>Contention Window</i>	64 ms
<i>DataPckt</i>	100 B	<i>MAC scheme cycle (IPCR-SMAC)</i>	4544 ms
<i>ACK</i>	10 B	<i>Duty Cycle</i>	50%

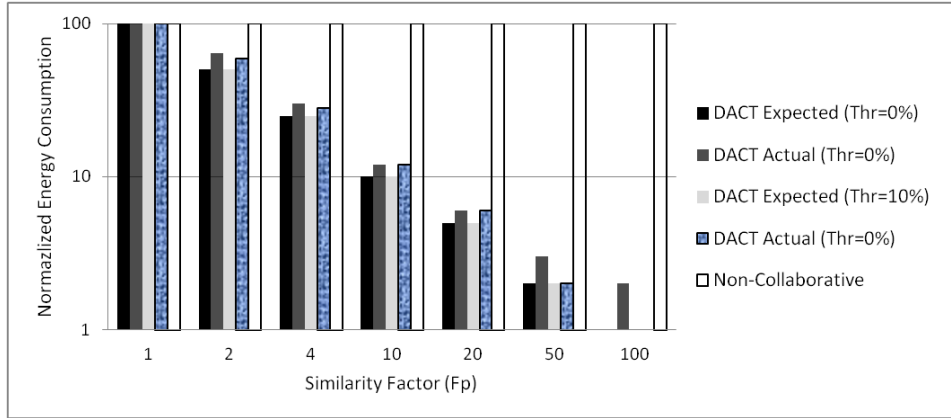
Since DACT gives each node in the network an equal opportunity of acquiring the channel, load balancing is implicit. Since it is the reporting node’s responsibility to assess its similarity to the data currently communicated over the channel, and the worst case scenario performance (i.e. no similarity between sensed data) will be comparable to that of conventional techniques. This greatly reduces the load required by the sink node, which is usually responsible for aggregating the unconditionally transmitted data (from reporting nodes), by distributing the effort over the entire neighborhood.

## 5 Evaluation

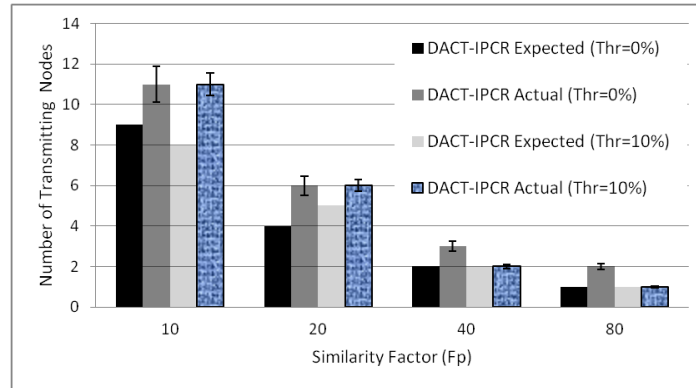
In this section, we evaluate DACT via ns-2 simulations [1]. Each node in our simulations has a single omni-directional antenna and follows ns-2’s commonly used combined free space and two-ray-ground reflection propagation model for wireless sensor networks. The underlying MAC scheme is Sensor MAC (SMAC) [18], and we assume that nodes follow a single sleep/wakeup schedule. The transmission range and carrier sensing range are modeling a 914MHz Lucent WaveLAN DSSS (Direct Sequence Spread Spectrum) radio interface which was used in several previous studies [8, 16]. Although this radio is not typical for a low power WSN node, but we use its parameters to make our results comparable to those reported in previous work [8, 16]. Furthermore, measurements have shown that similar proportions of the carrier sensing range to the transmission range are observed in some nodes [3, 16].

We test DACT on a 100 node randomly deployed electro-magnetic radiation sensor network, where sensors are deployed following a uniform random distribution, covering a total area of 10,000 square feet. The network has a randomly selected sink to which all nodes are required to report. All nodes are assumed to be in each other’s communication range (forming a single neighborhood). One way of extending this to multiple neighborhoods is to have border nodes of each neighborhood follow multiple schedules as in SMAC [18]. Each simulation is an average of 15 runs, each lasting for 7000 seconds. Key network simulation parameters are summarized in Table 2.

In order to accurately study DACT’s performance, we need to minimize the influence of the underlying MAC scheme for our simulation results. To do so, we keep the duty cycle relatively high (at 50% ) and the data rate relatively low



**Fig. 1.** Comparing DACT to non-collaborative communication solutions on a 100-node EH-WSN



**Fig. 2.** Comparing Analytical and Simulation Results of IPCR on DACT for the 100-node EH-WSN (Figure reproduced based on information from [10])

(at 1 packet every 50 seconds). These values ensure that the underlying MAC protocol (IPCR in our case) is operating under relaxed conditions and allows accurate evaluation of DACT strategy.

Figure 1 shows the normalized energy consumption for different similarity factors at different threshold levels. When the threshold is 0% (set by the consumer indicating maximum accuracy requirement) as shown in Fig. 1, the average energy consumed is at its maximum. This is expected due to maximum accuracy requirement. Since techniques that rely solely on spatial and/or temporal field correlations in their operation (e.g. [11, 17]) do not realize field similarity occurring only in the data space, their message complexity is not affected by uncorrelated field similarity and remains at its maximum. Figure 1 also reflects the result of increasing the threshold to 10%, which decreases the average number

of reporting nodes by approximately 10% , which is directly reflected in total energy consumption

In Fig. 2, we take a closer look at the effect of IPCR on DACT for larger similarity factor  $F_p$ . Our experimental results show reduction in the communication complexity and is in agreement with the theoretical analysis in section 4.3. Minor deviation from the average is also shown in the figure using error bars.

## 6 Conclusion

In this paper, we explored potential gains from monitoring energy consumption in real-time by utilizing in-network collaborative information processing to reduce information redundancy and communication operations in Energy Harvesting Wireless Sensor Networks (EH-WSN). We introduced a Data Aware Communication Technique (DACT), which exploits similarities in the sensory data to reduce communication redundancy via a clever similarity assessment technique. DACT is designed without any spatial or temporal assumptions about the data and can be applied to any energy monitoring and reduction application. DACT detects such similarities in the data space and takes advantage of them to reduce communication complexity and hence reduce energy consumption by up to 90%.

## References

- [1] UC Berkeley, LBL, USC/ISI and Xerox Parc, NS-2 documentation and software, Version 2.29, October 2005, <http://www.isi.edu/nsnam/ns>
- [2] Aledhari, M., Saeed, F.: Design and implementation of network transfer protocol for big genomic data. In: IEEE International Congress on Big Data (IEEE Big Data Congress 2015). IEEE (2015)
- [3] Anastasi, G., Falchi, A., Passarella, A., Conti, M., Gregori, E.: Performance measurements of motes sensor networks. In: Proceedings of the ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems. pp. 174–181. MSWiM '04 (2004)
- [4] Bajwa, R., Rajagopal, R., Varaiya, P., Kavalier, R.: In-pavement wireless sensor network for vehicle classification. In: Proceedings of the IEEE International Conference on Information Processing in Sensor Networks (IPSN), 2011. pp. 85–96
- [5] Blum, R., Sadler, B.: Energy efficient signal detection in sensor networks using ordered transmissions. *Signal Processing, IEEE Transactions on* 56(7), 3229–3235 (2008)
- [6] Cano, C., Bellalta, B., Sfairopoulou, A., Oliver, M., Barcel, J.: Taking advantage of overhearing in low power listening wsns: A performance analysis of the lwt-mac protocol. *Mobile Networks and Applications* 16, 613–628 (2011), <http://dx.doi.org/10.1007/s11036-010-0280-4>, 10.1007/s11036-010-0280-4
- [7] Chen, Z., Khokhar, A.: Self organization and energy efficient TDMA MAC protocol by wake up for wireless sensor networks. In: IEEE SECON. vol. 2004, pp. 210–207 (2004)
- [8] Du, S., Saha, A., Johnson, D.: RMAC: A routing-enhanced duty-cycle MAC protocol for wireless sensor networks. In: Proceedings of the IEEE International Conference on Computer Communications (INFOCOM). pp. 1478–1486 (May 2007)

- [9] Hefeida, M., Khokhar, A.: A cross-layer approach for context-aware data gathering in wireless sensor networks. In: Global Communications Conference (GLOBECOM), 2012 IEEE. pp. 238–243 (2012)
- [10] Hefeida, M., Khokhar, A.: Energy conservation in wsns: A collaborative information processing approach. In: Wireless Communications and Mobile Computing Conference (IWCMC), 2014 International. pp. 1118–1123 (Aug 2014)
- [11] Hesham, L., Sultan, A., Nafie, M., Digham, F.: Cooperative sensing with sequential ordered transmissions to secondary fusion center. In: Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on. pp. 2988–2991 (may 2011)
- [12] Iima, Y., Kanzaki, A., Hara, T., Nishio, S.: Overhearing-based data transmission reduction for periodical data gathering in wireless sensor networks. In: Complex, Intelligent and Software Intensive Systems, 2009. CISIS '09. International Conference on. pp. 1048–1053 (march 2009)
- [13] Lee, V.C.: Energy Harvesting for Wireless Sensor Network. Ph.D. thesis, University of California, Berkeley (2012)
- [14] Lin, X., Kwok, Y., Wang, H.: Cross-layer design for energy efficient communication in wireless sensor networks. *Wireless Communications and Mobile Computing* 9(2), 251–268 (2009)
- [15] Sartipi, M., Fletcher, R.: Energy-efficient data acquisition in wireless sensor networks using compressed sensing. In: 2011 Data Compression Conference. pp. 223–232. IEEE (2011)
- [16] Sun, Y., Du, S., Gurewitz, O., Johnson, D.: DW-MAC: a low latency, energy efficient demand-wakeup MAC protocol for wireless sensor networks. In: Proceedings of the ACM international symposium on Mobile ad hoc networking and computing (MobiHoc). pp. 53–62 (2008)
- [17] Vuran, M., Akyildiz, I.: Spatial correlation-based collaborative medium access control in wireless sensor networks. *IEEE/ACM Transactions on Networking (TON)* 14(2), 329 (2006)
- [18] Ye, W., Heidemann, J., Estrin, D.: An energy-efficient mac protocol for wireless sensor networks. In: Proceedings of the IEEE INFOCOM. vol. 3, pp. 1567–1576 (2002)
- [19] Ye, W., Heidemann, J., Estrin, D.: Medium access control with coordinated adaptive sleeping for wireless sensor networks. *IEEE/ACM Transactions on Networking* 12(3), 493–506 (2004)
- [20] Younis, O., Krunz, M., Ramasubramanian, S.: Node clustering in wireless sensor networks: recent developments and deployment challenges. *Network, IEEE* 20(3), 20–25 (2006)
- [21] Yu, L., Wang, N., Meng, X.: Real-time forest fire detection with wireless sensor networks. In: Wireless Communications, Networking and Mobile Computing, 2005. Proceedings. 2005 International Conference on. vol. 2, pp. 1214–1217. Ieee (2005)
- [22] Zou, Q., Zheng, S., Sayed, A.: Cooperative spectrum sensing via sequential detection for cognitive radio networks. In: Signal Processing Advances in Wireless Communications, 2009. SPAWC'09. IEEE 10th Workshop on. pp. 121–125. IEEE (2009)