Low-Energy Adaptive Clustering Hierarchy Using Affinity Propagation for Wireless Sensor Networks

Illsoo Sohn *Member, IEEE*, Jong-Ho Lee *Member, IEEE*, and Sang Hyun Lee *Member, IEEE*

***Abstract*—This letter proposes a new low-energy adaptive clustering hierarchy (LEACH) protocol for wireless sensor networks that use a distributed cluster formation based on affinity propagation (AP). The proposed LEACH protocol (LEACH-AP) enables a fully distributed control and resolves practical limitations of conventional LEACH-based protocols by simplifying network functionalities and reducing sensor hardware costs. Simulation results show that the proposed protocol outperforms existing LEACH-based protocols considerably in terms of network lifetime, energy dissipation rate, and total number of transferred bits.**

I. Introduction

Recent advances in battery-powered wireless sensors have enlarged their applications, including environmental monitoring, machine failure detection, surveillance, and internet-of-things applications [1]. Low-cost and small-sized wireless sensors have gained particular interest in efficient monitoring that involves thousands of wireless sensors in the measurement and report within a target area. Wireless sensors are typically scattered in a wide region without a sophisticated coordination. Since recharging the battery is almost impossible, wireless sensor networks (WSNs) are subject to energy management for maximizing their lifetime.

The low-energy adaptive clustering hierarchy (LEACH) protocol is a pioneering work in this type of applications [2], [3]. The LEACH protocol forms multiple clusters of nodes and designates a single cluster head (CH) node in each cluster, with the objective of minimizing the energy consumption of WSNs. In this hierarchy, CH nodes are responsible for the collection of the measurement from member nodes and the delivery of the aggregated information to the base station (BS) as illustrated in Fig. 1. Therefore, efficient formation of clusters is the crux of the LEACH protocol. A centralized version of LEACH (LEACH-C), where the BS is in charge of the cluster formation using all collected information, provides substantial improvement over distributed versions in terms of the network lifetime and data rate [2].

Recent studies have developed several enhanced variations of LEACH-C. In [4], each CH node avoids the energy depletion in its roles by assigning a vice-CH node in its cluster. LEACH-CE [5] considers the remaining energy information of the candidate CH node more actively. Upon the formation

I. Sohn and J.-H. Lee are with the Department of Electronic Engineering, Gachon University, Korea. S. H. Lee is with the Department of Electrical Engineering, Sejong University, Korea. (email: sanghyunlee@sejong.ac.kr).
This work was supported by the Gachon University research fund of 2014 (GCU-2014-0117). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. NRF-2015R1C1A1A01052529).

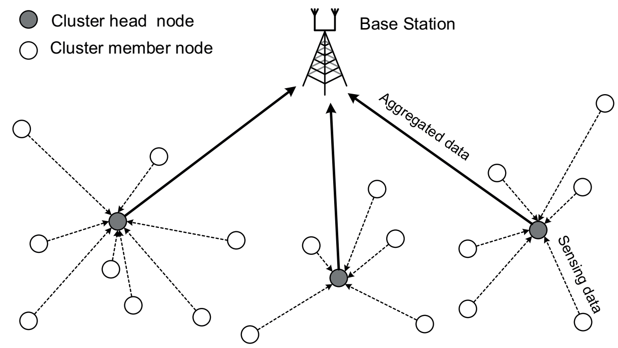


Fig. 1. A WSN model with low-energy adaptive clustering hierarchy.

of clusters, such a node with the largest remaining energy in an individual cluster becomes the CH node. This focuses on efficient distribution of the energy consumption over the WSN. LEACH-CKM [6] employs a sophisticated clustering technique instead of simulated annealing [7]. It uses K-means clustering which has been known to be a very efficient clustering technique. Nevertheless, LEACH-C and its variations have critical limitations in practical deployment: (i) Those protocols need knowledge of exact sensor locations obtained with additional hardware functionality such as GPS. (ii) The collection of the information required for clustering at the BS induces unnecessary signaling overheads. (iii) Clustering algorithms used in those protocols consider the energy consumption of links between the CH and member nodes only but fail to incorporate that of links between CHs and the BS, which has a significant impact on the cluster formation. In addition, those protocols rely on mean-based clustering where the center of a cluster is not necessarily a member. The resulting CH nodes chosen among nearby members may not be the most efficient. (iv) The optimal cluster number is very difficult to find for temporal variations in network topology. Fixing it to positive number K in existing protocols is obviously suboptimal.

This letter develops a new distributed energy-minimizing cluster formation strategy that resolves aforementioned drawbacks of existing LEACH-based protocols. The main idea is to formulate the WSN setup into median-based clustering, which guarantees the center of a cluster to be a member of the cluster, and to learn the optimal cluster number adaptively according to network topology. To this two-fold target, affinity propagation (AP) [8], a state-of-the-art message-passing-based clustering technique [9], is modified to handle the cluster formation. AP proves very efficient in handling various wireless network optimization tasks, such as forming primary-secondary user pairs in cognitive radio networks [10] and identifying the BS energy-saving status in green cellular networks [11]. The proposed LEACH protocol using AP (LEACH-AP) determines the cluster formation via iterative exchanges of simple messages among neighboring nodes, enabling a fully distributed control.

The rest of this letter is organized as follows. Section II describes system models, Section III proposes the improvement of LEACH using AP. After Section IV provides simulation results, Section V draws the conclusion.

II. System Model

Fig. 1 depicts a WSN model with a LEACH protocol. We consider N nodes scattered in a sensing region. A BS is located apart from the region and collects the measurement from the nodes. To establish a WSN, each node broadcasts a "Hello" message which contains its identity and energy status with a predefined transmit power. With the received power of Hello messages and the prior knowledge of the transmit power, nodes can estimate the distance among them. Based on distance profiles, nodes create a list of $\mathcal{N}$ dominant neighbors. Note that the location information is unnecessary unlike conventional protocols where the BS is responsible for finding distance profiles among nodes using GPS.

Fig. 2 depicts a basic LEACH protocol [2]. The transmission is repeated in series of rounds, each of which consists of set-up and steady-state phases. In the set-up phase, clusters are formed such that a subset of nodes become the corresponding CH nodes whereas the other nodes are members of those clusters. In the steady-state phase, the CH nodes schedule the transmission of member nodes while continuing the cluster formation and CH assignments. Upon receipt of the data from all member nodes, each CH node transmits the aggregated data to the BS at the end of each data frame.

We consider a packet-based energy consumption model [2], [12]. The energy consumption of the WSN consists of two parts: (i) digital-processing energy dissipation of electric circuitry which is proportional to the amount of the processed data and (ii) transmission energy dissipation of power amplifier which depends on both the amount of processed data and propagation distance. Let E_{elec}, ϵ_{fs}, ϵ_{mp}, d_0, E_{da}, and n denote the digital-processing energy per bit, the transmission energy per bit in free-space propagation, the transmission energy per bit in multi-path propagation, the reference distance for selecting between two propagation types, the energy consumption for data aggregation and the size of a cluster, respectively. In addition, let $\Delta(i,j)$ be an indicator function that yields one if $i \neq j$ and zero otherwise. The energy consumption required to transmit b bits at distance d is given by $E_{\text{TX}}(b,d) = bE_{\text{elec}} + b\epsilon_{\text{fs}}d^2$, if $d < d_0$, and $E_{\text{TX}}(b,d) = bE_{\text{elec}} + b\epsilon_{\text{mp}}d^4$, otherwise. Also, the energy consumption required to receive b bits is given by $E_{\text{RX}}(b) = bE_{\text{elec}}$. Presumably, the distance between the BS and the CH is larger than d_0, and the distance between the CH and member nodes is smaller than d_0 [2]. Therefore, the energy consumptions of CH and member nodes are given by $E_{\text{CH}} = (n-1)bE_{\text{elec}} + nbE_{\text{da}} + bE_{\text{elec}} + b\epsilon_{\text{mp}}d^4$ and $E_{\text{mem}} = bE_{\text{elec}} + b\epsilon_{\text{fs}}d^2$, respectively.

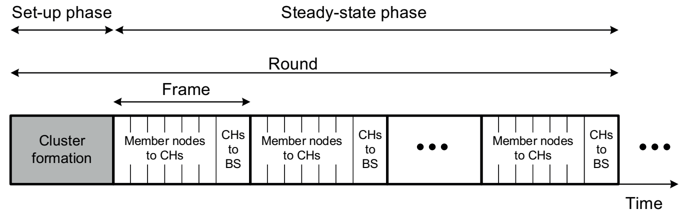


Fig. 2. A basic LEACH protocol.

III. LEACH-AP Protocol

A. Formulation

We formulate an optimization problem of the energy-minimizing clustering formation. For concreteness, integer variable c_i is introduced to represent the index of the CH node for node i. We denote the set of those indices by $\mathbf{c} = \{c_1, c_2, \dots, c_N\}$. Note that $c_i = j$ means that node j is the CH node of node i and node i is contained in the cluster associated with node j. If $c_i \neq i$, the expected energy consumption between node i and node c_i is $e(i,c_i) = E_{\text{mem}} = bE_{\text{elec}} + b\epsilon_{\text{fs}}d^2_{i,c_i}$. By contrast, $c_i = i$ implies that node i becomes a CH node. The expected energy consumption of node i results in $e(i,i) = (n_i-1)bE_{\text{elec}} + n_i bE_{\text{da}} + bE_{\text{elec}} + b\epsilon_{\text{mp}}d^4_{i,\text{BS}} \approx b\epsilon_{\text{mp}}d^4_{i,\text{BS}}$ since the BS is typically away from the WSN.

To formulate a clustering problem that maximizes the total similarities within clusters, the similarity between two nodes is defined as the negative energy consumption of the link between those nodes. The pairwise similarity of node pair (i,c_i) is defined as $s(i,c_i) = -e(i,c_i)\Delta(i,c_i)$. A large value of the similarity implies that those nodes are highly likely to belong to the same cluster. Let $\mathcal{E}_i$ and $\bar{\mathcal{E}}$ denote the remaining energy of node i and average remaining energy of active nodes, respectively. The self-similarity of node i is defined as $p(i) = -e(i,i)$, if $\mathcal{E}_i \geq \bar{\mathcal{E}}$, and $p(i) = -\infty$, otherwise. A large self-similarity of node i leads to high likelihood that node i becomes the CH node of a new cluster. Thus, nodes near the BS prefer the CH operation. The self-similarity set to $-\infty$ leads to handing the burden of the CH operation over to other nodes in case of low remaining energy, thereby preventing early energy depletion. In fact, the number of the clusters is learned from the relationship among self-similarities. We denote a constraint function of clustering operation by $\delta_k(\mathbf{c})$, which enforces a valid clustering formation ensuring that no CH node is a member of other clusters and that the cluster head is a node in the WSN. The resulting function is given by

$$\delta_k(\mathbf{c}) = \begin{cases} -\infty & \text{if } c_k \neq k \text{ but } \exists i : c_i = k, \\ p(k) & \text{else if } c_k = k, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Using the above definitions, the original clustering problem can be represented in an equivalent maximization problem as

$$\underset{\mathbf{c}}{\text{maximize}} \sum_{i=1}^{N} s(i,c_i) + \sum_{k=1}^{N} \delta_k(\mathbf{c}), \quad (2)$$

where the first term and the second term correspond to the energy consumption of member nodes and CH nodes,

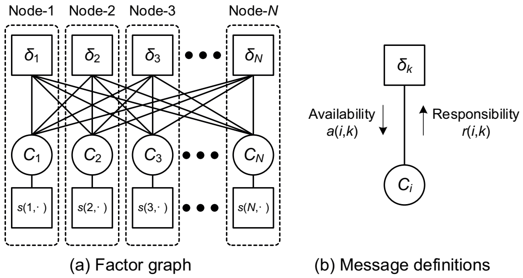


Fig. 3. The factor graph representation of cluster formation.

TABLE I
SIMULATION PARAMETERS

Description	Value
Digital processing energy E_{elec}	50nJ/bit
Transmission energy in free-space propagation ϵ_{fs}	10pJ/bit/m^2
Transmission energy in multi-path propagation ϵ_{mp}	0.0013pJ/bit/m^4
Energy consumption for data aggregation E_{da}	5nJ/bit/signal
Initial node energy	2J
Reporting interval	one hour
Cluster reformation interval	one week
Packet size (/node/report)	200bytes
Maximally allowed number of iterations $t_{\max}$	20
Stopping condition σ	0.001

respectively. Note that this formulation considers the energy consumption of the transmission from CH nodes to the BS, which existing approaches have not been able to handle.

B. Distributed Cluster Formation

We present a distributed algorithm for the optimization formulation in (2). It is convenient to represent the problem in a factor graph [9]. The factor graph is a visual representation for the problem, where a circle, a box, and an interconnecting edge denote an unknown CH node index, a constraint function, and their relationship, respectively (See Fig. 3(a)). Each node is connected with its $\mathcal{N}$ dominant neighbors based on distance profiles. Since messages are transferred in both directions of each edge, the factor graph defines the overall message flows of the distributed algorithm (See Fig. 3(b)). The solution is found by transferring messages along the edges iteratively. Meanwhile, a message associated with node i is obtained by considering the best valid configuration with c_i fixed to a certain node. The message transferred from function i to variable k, called *availability*, is computed as in [8] by

$$a(i,k) = \begin{cases} \Big[r(k,k) + \sum\limits_{i' \notin \{i,k\}} [r(i',k)]_+\Big]_- & \text{if } i \neq k, \\ \sum\limits_{i' \notin \{i,k\}} [r(i',k)]_+ & \text{if } i = k, \end{cases} \tag{3}$$

where $[\cdot]_+ = \max(0,\cdot)$ and $[\cdot]_- = \min(0,\cdot)$. The availability $a(i,k)$ sent from node k reflects the evidence for how appropriate it would be for node k to be a CH of node i. Also, the message transferred from variable k to function i, called *responsibility*, is derived in a similar way by

$$r(i,k) = \begin{cases} s(i,k) - \max\limits_{k' \neq k}(a(i,k') + s(i,k')) & \text{if } i \neq k, \\ p(i) - \max\limits_{k' \neq k}(a(i,k') + s(i,k')) & \text{if } i = k. \end{cases} \tag{4}$$

The responsibility $r(i,k)$ sent from node i reflects the evidence of how well-suited node k is to be the CH of node i.

In (3) and (4), the messages are updated using reciprocal previous values. The message transfers continue until message changes between consecutive iterations decay below a predetermined threshold σ or the iteration count reaches the number of maximally allowed iterations $t_{\max}$. These are system design parameters that can steer a stopping condition of the algorithm. Small σ and large $t_{\max}$ improve the performance at the cost of increased computational complexity. Upon termination of message transfers, node i determines its cluster head using

$$\hat{c}_i = \arg\max_k \left(a(i,k) + r(i,k)\right). \tag{5}$$

If $\hat{c}_i = i$, node i becomes a CH node and forms a new cluster. Otherwise, node i chooses node $\hat{c}_i$ as its CH node. The overall algorithm is summarized in Algorithm 1.

Algorithm 1 Distributed cluster formation algorithm

Initialize $t \leftarrow 0$, $a(i,k) \leftarrow 0$, and $r(i,k) \leftarrow 0$ for all (i,k).
repeat
  Update $a(i,k)$ messages using (3) and send to neighbors.
  Update $r(i,k)$ messages using (4) and send to neighbors.
  Increase $t \leftarrow t+1$
until $|\Delta r(i,k)| < \sigma$ for all (i,k) or iteration count $> t_{\max}$.
Node i chooses $\hat{c}_i$ as its cluster head based on (5).

The advantages of the proposed algorithm are multifold. This enables fully distributed processing and does not impose significant burden to the network side. Furthermore, there is no overhead for control signaling related to the information collection and dissemination along with the lift of the cost for optimization solvers at the BS. This approach can also reduce hardware costs for sensors because no positioning feature, such as GPS, is necessary. The exemption from GPS feature leads to the improvement of form factors in design of the sensor in terms of physical dimensions, weight, and battery capacity. On the other hand, the computational burden is low at an individual node. At a single iteration of the message transfer, the calculations of (3) and (4) involve only operations as many as the number of its neighbors $\mathcal{N}$. Therefore, the overall required computation for each node is $O(\mathcal{N}t_{\max}) = o(Nt_{\max})$.

IV. SIMULATION RESULTS

We evaluate the performance of the proposed LEACH-AP protocol in comparison with conventional LEACH-based protocols. A WSN with 100 wireless sensors is considered for simulation. Wireless sensors are distributed randomly in a 100m×100m region, i.e., from (0,0) to (100,100), and the BS is located at (50,175). The simulation parameters are listed in Table I according to [2], [12]. For all cases except for LEACH-AP, K is set to 5. The results are obtained by averaging over 1000 independently random wireless-sensor drops.

Fig. 4(a) compares the numbers of active nodes. LEACH-CKM increases the network lifetime, defined as the time

TABLE II
ENERGY CONSUMPTION PER A RECEIVED BIT [NANO-JOULE]

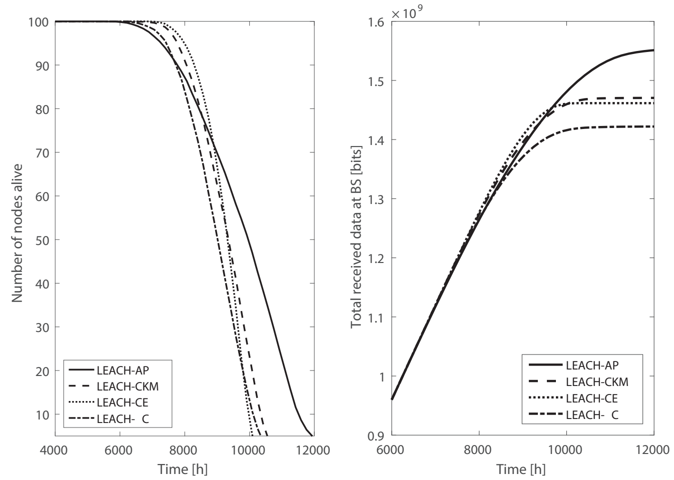
Case	LEACH-AP	LEACH-CKM	LEACH-CE	LEACH-C
A	1.29 (91.4%)	1.36 (96.7%)	1.37 (97.3%)	1.41 (100%)
B	2.66 (83.9%)	2.98 (94.0%)	3.12 (98.4%)	3.17 (100%)
C	1.73 (71.1%)	2.36 (96.7%)	2.41 (99.1%)	2.43 (100%)
D	2.89 (55.2%)	5.10 (97.4%)	5.04 (96.4%)	5.23 (100%)
E	1.30 (92.2%)	1.37 (97.7%)	1.37 (97.2%)	1.41 (100%)

during which at least K out of N nodes are alive [13], with the aid of an efficient clustering algorithm as compared with LEACH-C. LEACH-CE increases the first-node-death-time but rather decreases the network lifetime because its main target is to distribute energy consumption burden uniformly to all nodes. LEACH-AP extends the network lifetime by 15.5%, 18.3%, and 13.3% in comparison with LEACH-C, LEACH-CE, and LEACH-CKM, respectively. Fig. 4(b) shows the total received bits at the BS. The proposed protocol can deliver a considerably larger number of sensing data to the BS during the network lifetime than others, which is consistent with the previous results. Table II lists the energy consumption per a received bit in various scenarios. The variations from the initial setup of Case A are that the number of nodes decreases to $N = 50$ in Case B, that the BS is located farther at $(50, 250)$ in Case C, that the size of simulation region is doubled to 200m$\times$200m in Case D, and that the initial energy of a node is uniformly distributed within [1J, 3J] in Case E, respectively. In all scenarios, LEACH-AP outperforms the others consistently to a significant extent in terms of the energy consumption per a received bit. This noticeable performance improvement of LEACH-AP can be explained as follows. Fig. 5 shows how the number of clusters varies with time. At each instance, the proposed LEACH-AP changes the number of clusters adaptively if the resulting configuration reduces the overall energy consumption. This is a unique and attractive feature that renders the proposed approach very efficient for time-varying network topology. By contrast, all previous LEACH-based protocols are subject to a predetermined number of clusters and require a sophisticated optimization for the number of clusters with respect to the number of nodes, the size of region, and the location of CH nodes from the BS [2].

Fig. 6 shows the convergence property of LEACH-AP. The overall required computation for each node is only of $O(\mathcal{N}t_{\max})$. In general, the algorithm performance improves with the maximally allowed number of iterations. However, the performance improvement quickly becomes marginal after $20 - 30$ iterations for all cases. Thus, the required algorithm iterations for cluster formation can be kept small in practice. Since the cluster reformation interval is typically much longer than the data reporting interval, the energy consumption caused by computational costs and control signaling of the proposed algorithm can be expected to be minimal.

V. CONCLUSIONS

This letter addresses an enhancement of LEACH using AP for practical applications. The proposed LEACH-AP (i) does not require additional hardware functionality for location information, such as GPS, (ii) keeps the signaling overhead minimal regardless of network size, (iii) does not require the predetermination of the optimal number of clusters, and (iv) outperforms existing centralized approaches while providing aforementioned benefits inherited from its distributed nature.



(a) Number of nodes alive (b) Total number of received bits

Fig. 4. Number of nodes alive and total number of received bits versus time.

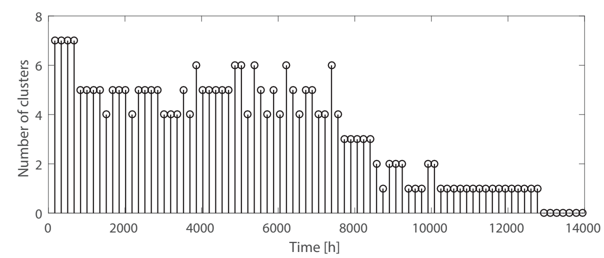


Fig. 5. Dynamic number of clusters versus time in LEACH-AP.

REFERENCES

[1] Z. Sheng, C. Mahapatra, C. Zhu, and V. C. M. Leung, "Recent advances in industrial wireless sensor networks toward efficient management in IoT," *IEEE Access*, vol. 3, no., pp. 622-637, 2015.
[2] W. B. Heinzelman *et al.*,"An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. Wirel. Commun.*, vol. 1, no. 4, pp. 660-670, Oct. 2002.
[3] X. Liu, "A survey on clustering routing protocols in wireless sensor networks," *Sensors*, vol. 12, no. 12, pp. 11113-11153, Aug. 2012.
[4] F. Zhao, Y. Xu, and R. Li, "Improved leach routing communication protocol for a wireless sensor network," *Int. J. Distr. Sensor Netw.*, vol. 2012, Article ID 649609, 6 pages, 2012.
[5] M. Tripathi *et al.*,"Energy Efficient Clustered Routing for Wireless Sensor Network," in *Proc. IEEE MSN'13*, pp. 330-335, 11-13 Dec. 2013.
[6] D. Mechta, S. Harous, I. Alem, and D. Khebbab, "LEACH-CKM: Low energy adaptive clustering hierarchy protocol with K-means and MTE," in *Proc. IIT'14*, Al Ain, pp. 99-103, 9-11 Nov. 2014.
[7] S. Z. Selim and K. Alsultan, "A simulated annealing algorithm for the clustering problem," *Patt. Recog.*, vol. 24, no. 10, pp. 1003-1008, 1991.
[8] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, 315, pp. 972-976, Feb. 2007.
[9] F. R. Kschischang *et al.*, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 498-519, Feb. 2001.
[10] M. Shamaiah, S. H. Lee, S. Vishwanath, and H. Vikalo, "Distributed algorithms for spectrum access in cognitive radio relay networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 10, pp. 1947-1957, Nov. 2012.
[11] S. H. Lee and I. Sohn, "Affinity propagation for energy-efficient BS operations in green cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4534-4545, Aug. 2015.

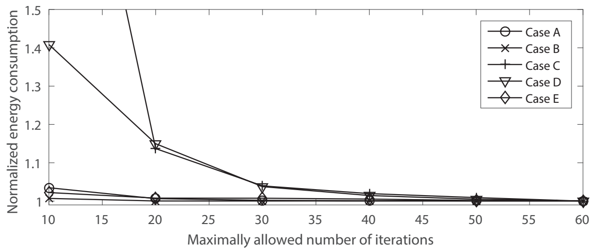


Fig. 6. Convergence property of LEACH-AP.

[12] J. Vazifehdan, R. V. Prasad, M. Jacobsson, and I. Niemegeers, "An analytical energy consumption model for packet transfer over wireless links," *IEEE Commun. Lett.*, vol. 16, no. 1, pp. 30-33, Jan. 2012.
[13] I. Dietrich and F. Dressler, "On the lifetime of wireless sensor networks," *ACM Trans. Sensor Networks*, vol. 5, no. 1, 2009.