Ad Hoc Networks 24 (2015) 172–185

Contents lists available at ScienceDirect

Ad Hoc Networks

journal homepage: www.elsevier.com/locate/adhoc

Towards prolonged lifetime for deployed WSNs in outdoor environment monitoring

Fadi M. Al-Turjman [a,*], Hossam S. Hassanein [b], Mohamad Ibnkahla [c]

[a] School of Engineering, University of Guelph, Guelph, Ontario N1G 2W1, Canada
[b] School of Computing, Queen's University, Kingston, Ontario K7L 3N6, Canada
[c] Department of Electrical and Computer Engineering, Queen's University, Kingston, Ontario K7L 3N6, Canada

ARTICLE INFO

Article history:
Received 22 February 2013
Received in revised form 5 July 2014
Accepted 26 August 2014
Available online 6 September 2014

Keywords:
Lifetime
Fault-tolerance
3D grid-based deployment
Wireless sensor networks

ABSTRACT

Recently, Wireless Sensor Networks (WSNs) emerged as a powerful and cost-efficient solution for unattended Outdoor Environment Monitoring (OEM) applications. These applications impose certain challenges on WSN deployment, including 3-Dimensional (3-D) settings, harsh operational conditions, and limited energy resources. To prolong lifetime of the deployed WSN, while mitigating the effects of these challenges, we propose the use of Relay Nodes (RNs) in addition to Sensor Nodes (SNs) in a distributed manner. While RNs facilitate reaching distant destinations, SNs can reserve their limited energy resources for sensing and data gathering. In addition, Mobile RNs (MRNs), which is a set of RNs capable of being reallocated (i.e. mobilized) at any point within the network lifetime, can be used to overcome possible link/node failure caused by the harsh conditions. It can also guarantee minimal energy consumption through imposing a balanced traffic distribution. This article proposes a 3-D grid-based deployment for heterogeneous WSNs (consisting of SNs, RNs, and MRNs). The problem is cast as a Mixed Integer Linear Program (MILP) optimization problem with the objective of maximizing the network lifetime while maintaining certain levels of fault-tolerance and cost-efficiency. Moreover, an Upper Bound (UB) on the deployed WSN lifetime, given that there are no unexpected node/link failures, has been driven. Based on practical/harsh experimental settings in OEM, intensive simulations show that the proposed grid-based deployment scheme can achieve an average of the expected UB.

1. Introduction

Wireless networking and advanced sensing technology have enabled the development of low-cost, power-efficient WSNs that can be used in various application domains such as healthcare, military, and OEM [2]. The main building block of a WSN is SNs. These nodes collect information, i.e. sense, some physical and/or chemical properties of a monitored environment and transmit their measurements to a central node known as the Base Station (BS). This transmission can either be periodic or on-demand [1]. Amongst the various application domains of WSNs, OEM attracted considerable attention due to its unique characteristics [3,4]. These include large monitored areas, isolated and distant territories, harsh operational conditions, and high probabilities of node and link failures [5,6]. Fortunately, a well-planned WSN deployment in such environments offers a reliable, yet cheap, means of decentralized data collection with minimal human intervention. However, in order to maintain a prolonged and reliable monitoring, the WSN needs to withstand the harsh operational conditions of outdoor environments, like heavy rains, snowfalls, sandstorms, extreme temperature variations, etc. These conditions may cause a significant percentage

* Corresponding author.
E-mail addresses: fadi@uoguelph.ca (F.M. Al-Turjman), hossam@cs.queensu.ca (H.S. Hassanein), ibnkahla@queensu.ca (M. Ibnkahla).

http://dx.doi.org/10.1016/j.adhoc.2014.08.017
1570-8705/

of node and link failure [7,8], which can be captured and statistically quantified using the Probability of Node Failure (PNF) and Probability of Disconnected Nodes (PDN). Hence, a well-planned WSN deployment should reduce these two probabilities [9]. While PNF can partially be reduced through proper placement and packaging of the nodes; PDN can significantly be reduced by a fault-tolerant deployment [10]. Fault tolerance is a pivotal step for a sustained and reliable monitoring. It is achieved through injecting a certain level of redundancy in the network such that it can withstand a given percentage of failure while maintaining the desired monitoring level.

Due to the scarcity of energy resources in outdoor environments, network nodes are almost always battery-powered. However, since a WSN in an OEM application is envisioned to work unattended for long periods of time, a stringent constraint is imposed on the energy consumption per node. This becomes a serious challenge when the network monitoring area is huge. In this case, energy can be drastically consumed if a distant BS is to be reached directly by all the SNs with a blind knowledge of the remaining energy budget per node in the network. To overcome this, a distributed system of RNs/SNs can be used [15]. A RN is a dedicated communication node with larger energy storage, i.e. larger battery, than regular SNs, capable of collecting data from a cluster of SNs and passes it to the BS. RNs can either be static or mobile. Unlike Static RNs (SRNs) that are located once within the network lifetime, Mobile RNs (MRNs) are given certain mobility features such that they can be relocated on demand [16]. The use of this type of RNs helps resolving bottleneck problems during the network lifetime. In fact, MRNs can be seen as a proactive solution to maintain connectivity and fault-tolerance when some communication paths are running out of energy or losing connectivity [17]. Also, one of the unique features of OEM applications is the 3-D space monitoring where the height of a node is as important as its horizontal position [12], which cannot be considered by 2D deployment algorithms. For instance, in monitoring the gigantic redwood trees in California, some experiments required placing the sensors at varying heights ranging from the ground surface up to tens of meters [13]. Moreover, monitoring the intensity of certain gases, like CO_2 [28,15,14], requires sensor placement at different heights such that monitoring coverage and accuracy requirements are met. Such a 3-D monitoring can easily be secured if the coverage space is modeled as a 3-D grid, which is a typical coverage model in OEM applications [2,28]. Hence, network nodes can only be placed at the vertices of this grid. In fact, the grid model limits the search space to a finite number of points. The shape of the grid building units can either be a cube, a hexagonal, an octahedron or any regular shape chosen to meet certain coverage levels [19]. Other advantages of the 3-D grid modeling include exclusion of all positions where node deployment is not possible, and accurate description of the possible routing paths [18]. In spite of the aforementioned grid advantages, placing the WSN nodes on the grid vertices might affect their deployment optimality. Nevertheless, this effect is fortunately controllable by the grid edge length (i.e. the deployment optimality is proportional with the count of vertices, and inversely proportional with the grid edge length). Thus, a more restricted search space (without affecting the deployment optimality) is required. Meanwhile, the overall network cost is proportional to the total number of nodes deployed. Hence, the lower the number of nodes, the lower the overall cost. However, communication reliability and fault tolerance require abundant node deployment. Consequently, a tradeoff exists between the overall cost and the network performance. Hence, the network deployment problem can be best modeled as an optimization problem. The objective is to maximize the network lifetime through reducing the energy consumption, while the constraints are cost efficiency, communication reliability, and fault-tolerance. This problem will be mathematically modeled in subsequent sections.

1.1. Related work

Recently, there have been several proposals for an energy-efficient, lifetime-maximizing WSN deployment. These proposals differ both in the type of nodes used in the network and their deployment strategies. In terms of the type of nodes used, most of the work in literature has considered the use of only sensor nodes deployed in a target area. Very few researchers have considered the use of RNs to improve the communication range and network lifetime. PEDAP and its power-aware version, PEDAP-PA [31], L-PEDAP [32], EESR [33], AND MLDA [34] are examples of WSNs that consider only sensor nodes deployed in a target area, and implement various schemes to improve the network lifetime. For instance, PEDAP considers minimizing the total energy expended by the network in a round of communication, but it does not consider the issue of balancing the energy consumption among the nodes. It consumes less energy to find a route and is able to achieve a good lifetime for the last node, but does not provide for load balancing among SNs and reliable communication in the network. PEDAP-PA is an improvised version of PEDAP that considers balancing the energy consumption among the SNs by computing their remaining energy using a cost function. However, this cost function considers only the transmitting nodes' residual energy, and the routing tree is recomputed after a pre-defined number of rounds, which is a major drawback when considering an improvement in the reliability of the system. Moreover, both PEDAP and PEDAP-PA are centralized algorithms that were designed for smaller deployment areas and might be unsuitable for large scale deployments such as OEM applications. EESR and L-PEDAP consider the remaining energy levels at both the transmitter and transceiver SNs to achieve better load balancing. EESR uses Kruskal's algorithm for the routing tree construction and works best when the SN is in the same transmission range and can communicate directly with the sink. L-PEDAP is capable of automatically re-routing a packet to the destination when it finds that the energy level of a node is less than the threshold value. Thus L-PEDAP achieves both load-balancing and reliable communication as it is capable of identifying node failure and recovering from it, unlike EESR. However, L-PEDAP fails to minimize the energy consumption and communication time. This is true even in the case of MLDA that fails to achieve the desired tradeoff between communication delay and network lifetime in the system. Thus, none of these works in existing literature have been able to satisfactorily address the problem of maximizing network lifetime while reducing the energy consumption, and considering the constraints of cost efficiency, communication reliability, and fault-tolerance.

Moving on to node deployment strategies, they are mainly classified as random and deterministic strategies (see Table 1). In random deployments, nodes' positions can be chosen in a purely random deployment plan or, based on a weighted random deployment plan, where the distributed nodes' density is not uniform in the monitored areas. For instance, K. Xu et al. [20], studied random deployment of static RNs in a two dimensional (2-D) plane. The authors proposed an efficient deployment strategy that maximizes the network lifetime when all RNs reach the BS with a single hop only. Motivated by the weakness of the uniform random deployment; the authors proposed a weighted random deployment strategy with a gradually increasing density of nodes as the distance to the BS increases. This strategy compensates the number of RNs for the energy needed to reach the BS. Hence, monitoring reliability can be sustained for longer periods of time, i.e. network lifetime is maximized.

In contrast, deterministic deployments aim at deploying nodes exactly on specific, predefined locations. These deployments can be accomplished through *centralized* or *distributed* approaches (see Table 1). In centralized approaches, global information gathering is required to end up with the targeted nodes' positions. For the most part each node requires a complete knowledge of the whole network topology. However, in distributed approaches deployment decisions are made based solely on some local knowledge per node. For example, a deterministic deployment strategy for mobile data collecting nodes was proposed in [22], assuming centralized knowledge and decisions made at the Base Station (BS). These mobile nodes move along a set of predefined tracks in the sensing field. In the proposed deployment strategy, SNs were able to relay data in addition to their sensing duties. It was shown that using data collectors (mobile relays) extends the network lifetime compared to conventional WSNs using static SNs only. In fact, data collectors were used earlier in [23,24]. The network lifetime was divided into equal length time intervals, called rounds. The data collectors are relocated at the beginning of each round based on a centralized algorithm running at the Base Station. The objective was to minimize the aggregate consumed energy during one round. It was shown that the optimal locations according to this objective function remain optimal even when the objective becomes to minimize the maximum energy consumed per SN. It should be remarked that these two energy metrics are not suitable for finding the optimal locations of mobile nodes since the optimal solutions will not be functions of time, i.e. time-independent. The reason is that the maximum or aggregate energy consumed per round might not change with time, and hence, locations of the data collector will not change with time. Consequently, the locations calculated may be far from optimal. Despite the advantages of these proposals, the deployed networks were prone to network partitioning and/or communication loss due to lack of fault tolerance. In addition, these proposals are designed for 2D deployment problem, which is not the case in OEM applications. In OEM, a 3D deployment plan for environment sensors is a must to achieve the desired outdoor observations [12,13]. Moreover, ignoring candidate sensory positions in the 3D space can waste numerous opportunities in reducing energy consumptions based on closer locations and in achieving better connectivity performance. In [25], a fault-tolerant random deployment was proposed. In particular, the authors proposed a distributed deployment algorithm to achieve a desired level of fault tolerance for all-SNs WSN. The transmission power of every node is gradually increased until either the distance between two neighboring nodes exceeds a specific threshold or the maximum transmission power is reached. In this deployment, fault tolerance is achieved at the expense of added cost. Transmission power adaptation requires complex hardware that raises the per-node cost, hence increasing the overall network cost. In addition, power adaptation results in added energy consumption, hence causing lifetime reduction. In [26], another fault-tolerant WSN deployment was proposed. The authors considered the case where at least two disjoint paths exist between each pair of SNs. To achieve a desired level of fault tolerance, deterministic RN placement was used. The problem was formulated as an optimization problem. However, it turned to be NP-hard. Hence, a polynomial time approximating

Table 1
A comparison between various deployment proposals in the literature.

Reference	Considered performance metrics[a]				Deployment approach		
	Cost	Connectivity	Fault-tolerance	Lifetime	Type	Centralized/decentralized	Targeted space
[9]	√	√	–	–	Deterministic	Decentralized	2D
[11]	√	√	–	√	Deterministic	Decentralized	2D
[14]	–	√	√	–	Random	Centralized	2D
[20]	√	√	–	–	Random	Centralized	2D
[22]	√	√	–	√	Deterministic	Centralized	2D
[23]	–	√	–	√	Deterministic	Centralized	2D
[25]	–	√	√	–	Random	Centralized	3D
[26]	√	√	√	–	Deterministic (grid-based)	Centralized	2D
[27]	–	√	√	√	Deterministic	Centralized	2D
Our work	√	√	√	√	Deterministic (grid-based)	Decentralized	3D

[a] The √ in this column indicates that the corresponding metric is considered. The- means the metric is not considered.

algorithm was proposed instead. This algorithm identifies candidate positions for RNs that cover the maximum number of SNs while assuming a regular communication range shape. Thus, RNs positions may not be accurate since it solely depends on the transmission ranges that are irregular in practice. Alternatively, fault tolerance could be achieved through deploying spare (redundant) node. In fact, faulty RNs are even more harmful to the network than SNs. This behavior was shown in [3] for a deterministic, grid-based deployment. Hence, an efficient fault-tolerance should account for faulty SNs as well as RNs.

1.2. Paper contributions

In this article, a comprehensive network deployment problem is considered based on decentralized algorithms running not only at the system sink (BS), but also at the core network nodes (SNs/RNs). SNs and RNs are jointly deployed. In addition, some MRNs are used to release the pressure from overloaded paths and fix any connectivity problems. Thus, the targeted problem can be stated as follows: *Given a 3D deployment space and a limited number of SNs and static/mobile RNs, find the optimal positions that prolong the network lifetime*[1] *while maintaining connectivity & certain fault-tolerance constraints*. Accordingly, our main contributions towards solving this problem can be summarized as

1. We overcome the huge search space of the candidate RNs positions by finding a subset of the grid vertices for these RNs based on their intersecting communication ranges.
2. The optimization problem is divided into *initial deployment* and *periodic redeployment*. In the initial deployment, the optimal locations of all nodes are found. In the periodic deployment, MRNs are relocated based on a decentralized decision made by deployed nodes at their present positions.
3. Efficient energy metrics, the *minimum node residual energy* and the *total energy consumed*, are used to maximize the network lifetime. These two metrics guarantee an influential MRN relocation.
4. An upper bound for the maximum network lifetime in ideal operation conditions is derived. This bound is used to show the performance gains achievable by the proposed two-phase solution.

The remainder of this paper is organized as follows. Section 2 describes the system model and the mathematical framework. The proposed deployment strategy is presented and discussed in Section 3. Section 4 presents the numerical results, while conclusions are drawn in Section 5.

2. System models

In this section we describe the communication model, the network architecture, and the lifetime model used in this article. All three models were tailored to suit OEM applications.

[1] See Definition 1.

2.1. Communication model

In practice, the signal level at distance from a transmitter varies depending on the surrounding environment. These variations are captured through the so called log-normal shadowing model. According to this model, the signal level at distance from a transmitter follows a log-normal distribution centered on the average power value at that point [29]. Mathematically, this can be written as

$$\begin{aligned} P_d(d) &= P_s - P_{\text{loss}}(d) \\ &= P_s - P_{\text{loss}}(r_0) - 10n\log\left(\frac{r}{r_0}\right) + \chi, \end{aligned} \tag{1}$$

where P_s is the transmission power, $P_{\text{loss}}(r_0)$ is the path loss measured at reference distance r_0 from the transmitter, n is an environment dependent path loss exponent, and χ is a normally distributed random variable with zero mean and variance σ^2, i.e. $\chi \sim \mathcal{N}(0, \sigma^2)$. With the aid of this model, the probability of successful communication between two nodes separated with a distance r can be calculated as follows. Assume $P_{\min}$ is the minimum acceptable signal level for successful communication between a source S and a destination D separated by distance r. The probability of successful communication is $\rho[S, D] = Pr[P_d(r) \geqslant P_{\min}]$. After some mathematical manipulations, $\rho[S, D]$ can be written as

$$\rho[S, D] = Q\left(\frac{P_{\min} - P_s - P_{\text{loss}}(r_0) - 10n\log(r/r_0)}{\sigma}\right), \tag{2}$$

where $Q(\cdot)$ is the Q-function defined as $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}dt$. In this paper, the probability of successful communication between nodes i and j should exceed a certain threshold, τ_1. Hence, the condition $\rho[i, j] \geqslant \tau_1$ will be used. The percentage of successful connectivity τ_1 is a design parameter. In fact, if identical transceiver specifications are used for all nodes; choosing a particular τ_1 automatically specifies the maximum distance between any two directly communicating nodes.

2.2. Network model

Consider a WSN where nodes are logically grouped into two layers, a lower layer consisting of the SNs, and an upper layer consisting of all the RNs. SNs forward their sensing data to a neighboring RN in the upper layer. On the other hand, RNs communicate periodically with the BS, either directly or via other RNs, to deliver the aggregated traffic from the SNs. Since these RNs have longer transmission range compared to Sensor Nodes (SNs), SNs in such a 2-tier architecture can invest their energy only in data gathering, and RNs can take care of communicating the gathered data to the BS. This helps to prolong the lifetime of the WSNs, which is very desired in OEM applications operating in harsh and energy-constrained environments.

In the grid-based network architecture, the grid-edge length is the transmission range r. Nodes can only be placed at the vertices of the 3-D grid such that the maximum number of Event Centers (ECs) is monitored. An EC is a location where the targeted phenomena can be monitored. The network topology is modeled as a graph $G = (V, E)$, where $V = \{n_0, n_1, \ldots, n_{v-1}\}$ is the set of v candidate grid vertices, E is the set of bidirectional links (edges) between the deployed nodes. Furthermore, the link between vertices i and j belongs to E if and only if the condition $\rho[i, j] \geqslant \tau_1$ is met. The network consists of Q_{SN} SNs and Q_{RN} RNs. If we let α_i^{SN} and α_i^{RN} be two binary variables such that $\alpha_i^{SN} = 1$ if a SN is placed at vertex i, $\alpha_i^{RN} = 1$ if a RN is placed at vertex i, and $\alpha_i^{SN} = 0, \alpha_i^{RN} = 0$ otherwise, then we can write

$$\sum_{i=1}^{v} \alpha_i^{SN} = Q_{SN}, \tag{3}$$

$$\sum_{i=1}^{v} \alpha_i^{RN} = Q_{RN}. \tag{4}$$

To achieve complete network coverage, the $Q_{SN} + Q_{RN}$ nodes should be distributed such that: (1) every EC is covered by at least one SN, (2) every SN is connected to at least one RN, and (3) every RN is connected to the BS either directly or indirectly via other RNs. These three requirements can mathematically be written as follows. Let $\beta_{i,\ j}$ be a binary variable whose value is 1 if the ith vertex is a candidate position to sense the jth EC and *0* otherwise. Consequently, the first requirement can be written as

$$\sum_{i=1}^{v} \alpha_i^{SN} \cdot \beta_{ij} \geqslant 1, \quad \forall j \in S_{EC}, \tag{5}$$

where S_{EC} is the set of ECs. To guarantee the communication between the lower layer and the upper layer in the network, the second requirement can be written as

$$\sum_{i=1}^{v} \alpha_i^{SN} . \alpha_j^{RN} \geqslant 1, \quad \forall i \in V \ \&\ j \in N(i), \tag{6}$$

where $N(i)$ is a set of neighboring indices such that $j \in N(i)$ if the jth vertex is within the transmission range of the ith node, i.e. $\rho[i, j] \geqslant \tau_1$. Finally, to guarantee that every RN can reach the BS either directly (one hop) or indirectly (multiple hops), the third requirement can be written as

$$\alpha_j^{RN} \cdot \left(\sum_{i \in \{N(BS), M(N(BS))\}} \alpha_i^{RN} \right) \geqslant 1, \quad \forall j \notin N(BS), \tag{7}$$

where $j \in M(N(BS))$ if the jth node can reach the BS either directly or indirectly.

2.3. Lifetime & energy models

Due to the harshness of outdoor environments, nodes and communication links are prone to failure. Losing some nodes and links may isolate other functional nodes. This problem can be overcome by deploying redundant nodes. Hence, deployment of redundant nodes helps achieving fault-tolerance, and thus prolongs the network lifetime. This concept can formally be defined as follows:

Definition 1 (*Network Lifetime*). is the time span (in rounds) from network deployment to the instant when the percentage of alive[2] and connected irredundant SNs and RNs falls below a specific threshold τ_2.

Notice that the remaining nodes, in addition to being alive, need to be connected to the BS either directly or indirectly. In order to measure the network lifetime, a measuring unit needs to be defined. In this work, we adopt the concept of a round as the lifetime metric. A round is the time span t_{round} over which every EC reports to the BS at least once. At the end of every round, the total energy consumed by the ith node can be written as

$$E_{\text{cons}}^{\text{i}} = \sum_{\text{Per round}} J_{\text{tr}} + \sum_{\text{Per round}} J_{\text{rec}}, \tag{8}$$

where $J_{tr} = L(\varepsilon_1 + \varepsilon_2 d^n)$ is the energy consumed for transmitting a data packet of length L to a receiver located r meters from the transmitter. Similarly, $J_{rec} = L\beta$ is the energy consumed for receiving a packet of the same length [21]. The parameters ε_1, ε_2, and β are hardware specific parameters of the used transceivers. Accordingly, if the initial energy of the ith node, E_{init}^i, is known, its remaining energy, E_{rem}^i, at the end of the round can readily be calculated as $E_{\text{rem}}^i = E_{\text{init}}^i - E_{\text{cons}}^i$. At the end of every round, the total energy consumed by all nodes can be written as $E_{\text{cons}}^{\text{tot}} = \sum_{Q_{SN}} E_{\text{cons}}^i + \sum_{Q_{RN}} E_{\text{cons}}^i$. Since all RNs may transmit and/or receive data in every round while all SNs transmit their measurements, the total energy consumed per round can be written as

$$E_{\text{cons}}^{\text{tot}} = \sum_{i=1}^{v} \alpha_i^{RN} \left(\sum_{j \in N(i)} J_{rec} f_{ij} + \sum_{j \in N(i)} J_{tr} f_{ij} \right) + \sum_{i=1}^{v} \alpha_i^{SN} \left(\sum_{j \in N(i)} J_{tr} f_{ij} \right), \tag{9}$$

where f_{ij} is the traffic from node i to j measured in bits per second (bps). The way Eqs. (3)–(9) are presented lends the network architecture and the energy model smoothly into the lifetime maximization problem to be discussed in the following section.

3. Deployment strategy

The deployment problem studied in this paper has an infinitely large search space. To limit this infinite search space to a manageable number of points, the 3-D grid model is used. The objective is to find the optimal locations of $Q_{SN} + Q_{RN}$ nodes among v grid vertices that maximize the network lifetime. The deployment strategy consists of two phases. The first phase aims at finding the optimal positions of all the nodes such that total energy consumption is minimized, and the second phase is launched at the end of every round to fix connectivity problem(s) and release the pressure from heavily loaded nodes. This two-phase deployment strategy is called *Optimized 3-D grid deployment (O3D).*

3.1. First phase of the O3D strategy

Let us start with the *Simplest Deployment Scheme (SDS)*. SDS aims at maximizing the network lifetime by finding: (1) the optimal deployment of all nodes in the network, and (2) the optimal routing paths from the SNs to the BS. This can be stated as follows: *what are the optimal deployment and routing strategies that need to be used to reduce the energy consumed per round*? While Eqs. (3)–(9) guarantee that a total of $Q_{SN} + Q_{RN}$ nodes satisfy the desired network topology; additional constraints are needed to control the routing paths. This can be done as follows. First, the traffic

[2] Alive nodes are those which *have enough energy for at least one more round.*

needs to be fairly divided amongst the deployed RNs to avoid node overload. Mathematically speaking, the conditions

$$\sum_{j\in N(i)} \alpha_i^{SN} \cdot f_{ij} \leqslant g_i^{SN}, \quad \forall i \in V, \tag{10}$$

$$\sum_{j\in N(i)} \alpha_i^{RN} \cdot f_{ij} - \sum_{k\in N(i)} \alpha_i^{RN} \cdot f_{ki} \leqslant g_i^{RN}, \quad \forall i \in V \tag{11}$$

need to be met. g_i^{SN} and g_i^{RN} are the generated traffic from the ith SN and RN, respectively, measured in bits per second (bps). Second, the limit of the available bandwidth for every node needs to be maintained. This can be achieved through the conditions

Algorithm 1. O3D First Phase Deployment.

Function OptIniDep (*IS*: Initial Set of ECs & BS coordinates to construct *N*, *V*)
Input:
A set *IS* of the ECs and BS coordinates.
A set *V* of the candidate grid vertices.
Output:
A set *N* of the SNs, SRNs, and BS coordinates.
Begin
Initialize: k-values, v, S_{EC}, g_i^{SN}, $g_i^{RN}, C_i^{SN}, C_i^{RN}, J_{tr}, J_{rec}, Q_{SN}, Q_{RN}$.
*PS*1 = Solve $\mathbb{P}_2$ in (18).
N= Set of coordinates of SNs, SRNs, and BS in *PS1*.
End

$$\sum_{j\in N(i)} \alpha_i^{SN} \cdot f_{ij} \leqslant C_i^{SN}, \quad \forall i \in V, \tag{12}$$

$$\sum_{j\in N(i)} \alpha_i^{RN} \cdot f_{ij} \leqslant C_i^{RN}, \quad \forall i \in V, \tag{13}$$

where C_i^{SN} and C_i^{RN} are the available bandwidths for the ith SN and RN, respectively, measured in bps. With the aid of these constraints, the SDS optimization problem can be summarized as follows $\mathbb{P}_1$:

$$\begin{array}{ll} \text{Minimize} & E_{\text{cons}}^{\text{tot}} \\ \text{Subject to} & \text{Eqs. } (3)-(9), \\ & \text{Eqs. } (10)-(13). \end{array} \tag{14}$$

Observe that we have intentionally divided the constraints into two groups, network architecture constraints and routing paths constraints. This division gives additional insights into the performance of the proposed scheme as shall be revealed in the results section. Despite that this scheme allows optimal deployment of a given set of nodes such that lifetime is maximized; the deployed nodes are prone to isolation and/or failure that render some ECs not covered. This is because no fault-tolerance constraint was imposed. To achieve this tolerance, the *Fault-tolerant Simplest Deployment Scheme (FSDS)* presents a fault-tolerant version of the SDS. Fault-tolerance is an energy consuming constraint that allows the network to withstand a certain level of faulty nodes while maintaining a desired level of coverage. Fault-tolerance can be quantified by the number or percentage of faulty nodes tolerated. When the number of non-operational nodes is less than k, the network still needs to recover the isolated ECs and nodes. This is achieved through the injection of redundant nodes as mentioned earlier. Consequently, data recovery can be defined as:

Definition 2. (*Data Recovery*). The existence of operational redundant nodes capable of covering isolated or partitioned ECs and nodes and routing their measurements to the BS.

With the aid of these definitions, we can summarize the objective of the FSDS scheme as follows: maximize the network lifetime through finding: (1) the optimal deployment of all nodes in the network, and (2) the optimal routing paths from the SNs to the BS, such that the deployed network is k fault-tolerant and data recovery is guaranteed. In other words, the FSDS scheme is an extension of the SDS with fault tolerance and data recovery constraints. Let us start with the fault-tolerance constraint. Every component in the network (EC, SN, or RN) shall be connected to more than one element in the upper level to achieve a desired level of redundancy. In other words, every EC needs to be covered by $k_1 \geqslant 1$ SNs, every SN needs to be connected to at least $k_2 \geqslant 1$ RNs, and every RN needs to reach the BS through at least $k_3 \geqslant 1$ routes. Consequently, (5)–(7) can be rewritten as:

$$\sum_{i=1}^{v} \alpha_i^{SN} \cdot \beta_{ij} \geqslant k_1, \quad \forall j \in S_{EC} \tag{15}$$

$$\sum_{i=1}^{v} \alpha_i^{SN} \cdot \alpha_j^{RN} \geqslant k_2, \quad \forall j \in V \ \& \ j \in N(i), \tag{16}$$

$$\alpha_j^{RN} \cdot \left(\sum_{i\in\{N(BS),M(N(BS))\}} \alpha_i^{RN} \right) \geqslant k_3, \quad \forall j \notin N(BS). \tag{17}$$

It should be mentioned here that the way (15)–(17) is written gives us flexibility in choosing the level of tolerance at all layers in the network. This allows more customized fault tolerance. However, to make the entire network a k fault tolerant network, we simply set $k_1 = k_2 = k_3 = k$. Finally, $\mathbb{P}_1$ can be rewritten with the modified constraints to get the FSDS in the form of $\mathbb{P}_2$:

$$\begin{array}{ll} \text{Minimize} & E_{\text{cons}}^{\text{tot}} \\ \text{Subject to} & \text{Eqs. } (3)-(9), \\ & \text{Eqs. } (10)-(13), \\ & \text{Eqs. } (15)-(17). \end{array} \tag{18}$$

By solving this optimization problem at the system sink (BS), a fault tolerant, maximized lifetime, WSN deployment will be achieved, which is also the first phase deployment of the O3D strategy summarized in Algorithm 1. An additional degree of freedom can be brought to the network through the use of MRNs. MRNs can be reallocated periodically such that a particular objective is achieved. In our case, we shall use it to maximize the network lifetime by reducing the energy consumption of a particular set of nodes that have been overloaded, as described next.

3.2. Second phase of the O3D strategy

After performing the first phase of the O3D in $\mathbb{P}_2$ which considers all the available RNs, i.e. Q_{RN}^S static RNs and Q_{RN}^M mobile RNs, $\left(Q_{RN} = Q_{RN}^S + Q_{RN}^M\right)$, a relocation of the MRNs is performed in the second phase of the O3D strategy, as shall subsequently be shown. However, since the second phase deployment will take place during the network run time, processing time is very critical and has to be very limited. Thus, searching all the grid vertices, v, in large-scale applications with large v like OEM applications, is a computationally expensive and time consuming process. This is due to the involved computational requirements for finding the network lifetime for a large number, $\binom{v - Q^S_{RN} - Q_{SN}}{Q^M_{RN}}$, of possible node locations. Therefore, a limited search space with reduced v is needed. Taking advantage of a distributed system running at the SNs and the static RNs to provide a previous knowledge of their positions to the BS, the MRNs may be placed on any grid vertex as long as they are within the probabilistic communication range of the largest number of SNs and/or static RNs. As a result, the search space is reduced while the accuracy of the deployment plan is not affected. To explain our method of finding this finite search space, we give the following two definitions.

Definition 3 (*Ideal Set*). A finite set of positions P is ideal if and only if it satisfies the following property: there exists an optimal placement of MRNs in which each relay is placed at a position in P.

We aim at finding such an ideal set in order to achieve more efficient discrete search space in which candidate positions do not include all the grid vertices but only a subset of it. This subset should have the highest potential to prolong the network lifetime and sustain its fault tolerance through maintaining the largest neighborhood. However, since computational complexity is proportional to the cardinality of P, a set with an even smaller cardinality is needed.

Definition 4 (*Candidate Grid Unit (CnGU)*). A candidate grid unit α is a grid unit that has a connected center with at least k_2 SN *or* k_3 static RN. The subset of SNs and static RNs coordinates connected to α is denoted by $C(\alpha)$.

The building unit of the 3-D grid is called a *Grid Unit (GU)*. A *GU* is said to be connected to a particular SN or static RN if the condition $\rho \geqslant \tau_1$ is met, where ρ is the probabilistic connectivity parameter between the *GU* center and that node.

Definition 5 (*Optimal Candidate Grid Unit (OCGU)*). A candidate GU α is optimal if there is no candidate GU β, where $C(\alpha) \subseteq C(\beta)$.

The OCGUs have the highest potential to place the MRNs based on Definitions 4 and 5. Accordingly; we have to show that an ideal set can be derived from the set of OCGUs. Towards this end, we state the following *Lemmas*.

Lemma 1. *For every CnGU β, there exist a OCGU α such that $C(\beta) \subseteq C(\alpha)$.*

Proof. If β is a OCGU, we choose α to be β itself. If β is not a OCGU then, by definition, there exists a CnGU α_1 such that $C(\beta) \subseteq C(\alpha_1)$. If α_1 is a OCGU, we choose β to be α_1, and if α_1 is not OCGU then, by definition, there exists another CnGU α_2 such that $C(\alpha_1) \subseteq C(\alpha_2)$. This process continues until a OCGU α_x is found; we choose α to be α_x. Thus, Lemma 1 holds. □

Lemma 2. *Finding a OCGU takes at most* $(n-1)$ *step, where* $n = Q_{SN} + Q^S_{RN}$.

Proof. By referring to the proof of Lemma 1, it is clear that $|C(\alpha_x)| \leqslant n$, and $|C(\alpha)| < |C(\alpha_1)| < |C(\alpha_2)| < \ldots < |C(\alpha_x)| \leqslant n$; where $|C|$ is the cardinality of C. Consequently, the process of finding the OCGU α_x takes a finite number of steps $\leqslant n-1$. □

Then, we introduce the following *Theorem*.

Theorem 1. *A set P that contains one position from every OCGU is ideal.*

Proof. To prove this *Theorem*, it is sufficient to show that for any arbitrary placement Z we can construct an equivalent[3] placement $\overline{Z}$ in which every MRN is placed at a position in P. To do so, assume that in Z, a MRN i is placed such that it is connected to a subset J of SNs/SRNs. It is obvious that there exists a CnGU β, such that $J \subseteq C(\beta)$. From Lemma 1, there exist a OCGU α such that $C(\beta) \subseteq C(\alpha)$. In $\overline{Z}$, we place i at the position in P that belongs to α, so that i is placed at a position in P and is still connected with all SNs/SRNs in J. By repeating for all MRNs, we construct a placement $\overline{Z}$ which is equivalent to Z, and thus Theorem 1 holds. □

In order to find all OCGUs, we need a data structure associated with each GU to store coordinates and total number of SNs/SRNs connected to the GU. We represent this data structure by the CnGU set $C(i)$, where i is the center of the CnGU. By computing $C(i)$, $\forall i \in V$, we can test whether a CnGU centered at i is optimal or not by searching for a set that has at least all elements of $C(i)$. In the following, Algorithm 2 is running independently at the SNs/RNs by the end of each round to collect residual energy and neighboring CnGUs per node. If any change occurs in the resultant gathered information, it will be broadcasted to update the system sink. Accordingly, Algorithms 3–6 will be executed at the system sink (BS). Algorithm 3 establishes the set of vertices that are within the communication range of a single GU based on the output of Algorithm 2; it runs in $O(n)$ time. Algorithm 4 tests whether a CnGU is optimal or not based on a local comparison between the resultant output of each SNs/RNs. Algorithm 5 uses Algorithms 3 and 4 to construct the ideal set P by finding all OCGUs. The overall complexity of Algorithm 5 is $O(n\log n)$.

Once we obtain the set P which contains one position (grid vertex coordinates) from each OCGU, the search space

[3] Equivalent in terms of connected SNs/SRNs. In other words, the placement of a MRN at position i, within the communication range of the nodes x and y, is equivalent to the placement of the same MRN node at position j within the communication range of the nodes x, y and z.

of the optimization problem to be formulated, subsequently, becomes much more limited. To start formulating the second phase optimization problem, we assume that an FSDS deployment has already been performed and that the network is running. Some nodes will be consuming more energy than others such that network bottlenecks start to appear.

Algorithm 2. Information Gathering.

Function InfoG (N)
Input:
  A set N of the GUs' coordinates.
Output:
  LoCGU: List of covered GUs by a SN/SRN j.
  E: remaining energy at SN/SRN j.
Begin
  If SN
   E = E^{SN}_{rem} **with reference to** Eq. (9)
  Endif
  If RN
   E = E^{RN}_{rem} **with reference to** Eq. (9)
  Endif
LoCGU(j):=∅;//list of covered GUs by node j
foreach GU center i **do**
    Compute $\rho[i, j]$;
    If $\rho[i, j] \geqslant \tau_1$
      LoCGU(j):=$i \cup$ *LoCGU*(j);
    endif
endfor
End

Algorithm 3. Creating Grid Units.

Function FindCandidateGridUnit (N, *LoCGUs*)
Input:
  N: set of the SNs and SRNs nodes' coordinates.
  LoCGUs: list of covered GUs by each SN/SRN.
Begin
foreach GU center i in *LoCGUs* **do**
  $C(i)$:=∅;
  x_i:=0; //where x_i represents number of nodes connected to the GU centered at i.//
  L_i:=∅; //L_i is the list of SN/SRN coordinates connected to the GU center i.//
  foreach SN/SRN j in *LoCGUs* do
    If i is covered by a SN/SRN
      $x_i = x_i + 1$;
      $L_i = L_i \cup j$;
    endif
  endfor
  If ($x_i \geqslant k$)
    $C(i)$:=$L_i \cup C(i)$;
  Endif
endfor
End

Algorithm 4. Testing whether $C(i)$ is Optimal or Not.

Function Optimal ($C(i)$, all non-empty grid unit sets)
Input:
  A set $C(i)$ for a specific grid unit center i & All non-empty sets of the grid units' centers.
Output:
  True if $C(i)$ is OCGU and False otherwise.
Begin
If $C(i)$:=∅ **do**
    return False;
endif
Search for a set $\overline{C}$ such that $C(i) \subseteq \overline{C}$.
If $\overline{C}$:= ∅ **do**
    return True;
else
    return False;
endif
End

Algorithm 5. Finding all OCGUs.

Function FindOCGUs (N: Set of SNs & SRNs)
Input:
  A set N of the SNs and SRNs nodes' coordinates.
Output:
  A set P that contains one position from every OCGU.
Begin
P:=∅;
FindCandidateGridUnit (N);
foreach $C(i)$ do
    If Optimal ($C(i)$, all non-empty grid unit sets) **do**
      P:={i} ∪ P;
    endif
endfor
End

To help these nodes, a total of Q^{M}_{RN} MRNs shall be relocated. The objective of this relocation process is two-fold: maximize the minimum residual energy amongst all nodes per round, and minimize the total energy consumption. Notice that the first part of the objective aims at maximizing a time-and-node-dependent value; hence relocation will always lead to new optimal locations whenever applied. The objective function can be written as

$$\text{Minimize} \quad E^{\text{tot}}_{\text{cons}} - E^{SN}_{\text{res}} - E^{RN}_{\text{res}} \tag{19}$$

Algorithm 6. O3D Second Phase Deployment.

Function MRNsP (N: constructed by SNs, SRNs & BS, P)
Input:
  A set N of the SNs, SRNs and BS nodes' coordinates.
  An ideal set $PS2$ of v candidate positions for the MRNs.

(*continued on next page*)

Output:
A set PS2 of the MRNs coordinates maximizing lifetime of N.
Begin
Initialize: k-values, $S_{EC}, g_i^{SN}, g_i^{RN}, C_i^{SN}, C_i^{RN}, Q_{RN}^{M}$, And make $V = P,\ v = |P|$.
$PS2$ = Solve $\mathbb{P}_3$ in (23).
End

where E_{res}^{SN} and E_{res}^{RN} are the minimum residual energy over all SNs and RNs, respectively, at the end of the round. Notice that RNs include both, static RNs and MRNs. To guarantee connectivity for the entire round, these two residual energies need to satisfy two conditions at the end of the round. The first one is

$$E_{\text{res}}^{SN}, E_{\text{res}}^{RN} \geqslant 0, \tag{20}$$

while the second one is

$$\begin{aligned} &\alpha_i^{SN} E_{\text{rem}}^{SN} - \sum_{j \in N(i)} \alpha_i^{SN} J_{tr} f_{ij} - \sum_{k \in N(i)} \alpha_i^{SN} J_{rec} f_{ki} \geqslant E_{\text{res}}^{SN}, \ \forall i \in V, \\ &\alpha_i^{RN} E_{\text{rem}}^{RN} - \sum_{j \in N(i)} \alpha_i^{RN} J_{tr} f_{ij} - \sum_{k \in N(i)} \alpha_i^{RN} J_{rec} f_{ki} \geqslant E_{\text{res}}^{RN}, \ \forall i \in V, \end{aligned} \tag{21}$$

where E_{rem}^{SN} and E_{rem}^{RN} are the remaining energies in the SNs and RNs, respectively. These two equations guarantee successful communication for at least one coming round. Furthermore, at the end of the relocation process, the number of MRNs needs be the same; whether relocated or kept in place. Hence, similar to (3) and (4), one can write

$$\sum_{i=1}^{v} \alpha_i^{MRN} = Q_{RN}^{M}. \tag{22}$$

Consequently, with the aid of these constraints, we can rephrase our optimization to be written as $\mathbb{P}_3$:

$$\begin{aligned} \text{Minimize} \quad & E_{\text{cons}}^{\text{tot}} - E_{\text{res}}^{SN} - E_{\text{res}}^{RN}, \\ \text{Subject to} \quad & \text{Eqs. } (3)-(9), \\ & \text{Eqs. } (10)-(13), \\ & \text{Eqs. } (15)-(17), \\ & \text{Eqs. } (20)-(22). \end{aligned} \tag{23}$$

By solving this optimization problem, a further fault tolerant and maximized lifetime WSN deployment will be achieved, which is also the second phase deployment of our O3D strategy summarized in Algorithm 6. In light of the large existing literature today on optimization formulations for lifetime in WSNs, it is worth pointing out that the two energy metrics (residual and consumed energy metrics) included in the objective function make the proposed MILP unique and particularly distinguished. As it makes the proposed approach more suitable for finding the optimal locations of mobile nodes since the optimal solutions will be a function of time, i.e. time-dependent. This is a typical use of the proposed mathematical model under realistic settings in mobile topologies.

Finally, based on the output of Algorithms 1 and 6, optimal locations of SNs and RNs in terms of maximum lifetime and limited cost budget are determined, in addition to fault-tolerance and data recovery constraints. This two-phase solution can easily be extended to consider other constraints such as coverage, data fidelity, as well as delay-tolerance through formulating and adding it to $\mathbb{P}_3$.

3.3. Lifetime theoretical analysis

In the previous section, we examined the placement problem for a heterogeneous WSN when both energy-efficient and fault-tolerant design factors are considered. However, once the network starts being operational deployed nodes start losing energy and facing harsh operational conditions that may lead to failures and increased risks of disconnection. Consequently, a second phase was proposed to reposition MRNs in order to overcome such conditions, and hence provide near optimal solutions in practical situations. As these conditions are scarcely predictable in practice, it is very difficult to predict what would be the maximum number of rounds a network can stay operational for. Therefore, we derive an Upper Bound (UB) on the number of rounds a WSN can spend during its lifetime, given that there are no unexpected node/link failures. Thereby, we assume the same notations in the system models section. Also, we define LT_{max} to be the maximum number of rounds a WSN can stay operational for, and $E_{min/r}^{SN}$ to be the minimum total energy consumed by SNs per round, and $E_{min/r}^{RN}$ to be the minimum total energy consumed by RNs per round. Assume E_{init}^{tot} is the initial total available energy before the network starts functioning, E_{init}^{SN} is the initial available energy per SN, and E_{init}^{RN} is the initial available energy per RN.

Theorem 2. *The lifetime of the deployed WSN is upper-bounded by:*

$$LT_{max} = \min\left\{ \frac{Q_{SN}.E_{init}^{SN}}{J_{tr}\left[\frac{Q_{SN}}{k_1}\sum_{i=1}^{\frac{Q_{SN}}{k_1}} g_i^{SN}\right]}, \frac{Q_{RN}.E_{init}^{RN}}{(J_{tr}+J_{rec})\left[\frac{Q_{RN}}{k_2}\sum_{i=1}^{\frac{Q_{RN}}{k_2}} g_i^{RN}\right]} \right\}. \tag{24}$$

Proof. As the minimum consumed energy per round by SNs is the required energy to deliver irredundant generated traffic (sensed data), the minimum energy consumed by these nodes per round is equal to the energy used in transmitting from irredundant SNs. Since the number of irredundant SNs is Q_{SN}/k_1; the minimum energy consumed per round is

$$E_{min/r}^{SN} = J_{tr}\left[\frac{Q_{SN}}{k_1}\sum_{i=1}^{\frac{Q_{SN}}{k_1}} g_i^{SN}\right]. \tag{25}$$

Similarly, the number of irredundant RNs is Q_{RN}/k_2; hence the minimum energy consumed per round is

$$E^{RN}_{min/r} = (J_{tr} + J_{rec})\left[\frac{Q_{RN}}{k_2}\sum_{i=1}^{\frac{Q_{RN}}{k_2}} g_i^{RN}\right]. \quad (26)$$

As the initial total available energy at SNs is equal to $Q_{SN}.E^{SN}_{init}$, the maximum number of rounds the SNs can stay operational for is $Q_{SN}.E^{SN}_{init}/E^{SN}_{min/r}$. Similarly, the maximum number of rounds the RNs can stay operational for is $Q_{RN}.E^{RN}_{init}/E^{RN}_{min/r}$. As the maximum number of rounds a WSN can stay operational for is controlled by the lifetime of SNs generating the sensed data, and the RNs relaying this data, the maximum number of rounds a WSN can stay operational for is

$$LT_{max} = min\left\{\frac{Q_{SN}.E^{SN}_{init}}{E^{SN}_{min/r}}, \frac{Q_{RN}.E^{RN}_{init}}{E^{RN}_{min/r}}\right\}. \quad (27)$$

By substituting (25) and (26) in (27), we achieve the lifetime UB described above in (24). □

It is worth mentioning that this UB depends only on the number of deployed nodes, initial node energy, node generation rate, redundancy level (i.e. k-value), and energy consumed for transmitting/receiving a packet. Consequently, to increase $LT_{\max}$ one can either increase the initial energy of the deployed nodes; decrease their energy consumption per packet, increase the number of nodes deployed, or increase the redundancy level. This UB is not only used in assessing the efficiency of our two-phase deployment strategy, but also any other deployment strategy aiming at maximizing the WSN lifetime.

4. Performance evaluation & discussion

In this section, we evaluate the performance of our proposed strategy in practical settings with different PNF and PDN conditions. We consider the SDS and FSDS schemes along with the UB as a baseline to the proposed O3D deployment strategy. In fact, simplified variations of SDS and FSDS schemes are widely studied in the literature [23,24]. To compare the performance of the three schemes, the following four performance metrics are used. The first metric is the average lifetime defined as the number of rounds through which the network operates. The second one is the average energy consumed per byte. This metric reflects the energy utilization efficiency. In fact, it relates energy consumption to network lifetime. The third metric is the *Ratio of Remaining Energy (RRE)*. RRE is the ratio of the total remaining energy in all nodes to their total initial energy. Finally, the fourth metric is the *Percentage of Packet Loss (PPL)*. PPL is the percentage of transmitted data packets that fails to reach the BS. It reflects the effects of bad communication channels and node failures. In studying these performance metrics, four parameters are used. These are the level of fault-tolerance k, the PNF, the PDN, and the ($Q_{SN} + Q_{RN}$) count.

4.1. Simulation model

The three deployment schemes: SDS, FSDS, and O3D, are applied based on experimental and realistic parameters that are typical in outdoor WSN applications [29,30]. Moreover, these schemes are executed on 500 randomly generated WSNs hierarchical graph topologies in order to get statistically stable results. The average results hold confidence intervals of no more than 2% of the average values at a 95% confidence level. For each topology, we apply a random node/link failure based on pre-specified PNF and PDN values, and performance metrics are computed accordingly. Dimensions of the monitored space are $900 \times 900 \times 300\ m^3$. We assume the same predefined fixed time schedule in [19] for traffic generation at the SNs and RNs, as it is a typical model for such hierarchical topologies in WSNs. Nodes positions are found by applying the three deployment strategies. We assume that each WSN is required to be operational for the maximum number of rounds using a maximum of 15 RNs and 1500 SNs (cost constraint). Based on real outdoor experimental measurements, the communication model parameters are set as shown in Table 2 [30]. The simulator determines whether or not a SN is connected to its neighbors according to the probabilistic communication model described earlier in Section 2.1. To simplify the presentation of the results, all the transmission ranges of SNs and RNs are assumed equal to 100 m. We use MATLAB *lp-solver v5.5* with a timeout of 15 min. In other words, the MILP of a particular round is solved during the last 15 min of the previous round. We remark that better lifetime performance can be achieved at the cost of more computational time complexity. However, this is not an issue in a typical OEM application, where a MILP can be left running for days without affecting the targeted application.

4.2. Simulation results

While the SDS strategy finds the optimal locations of the SNs and RNs in order to achieve the maximum network lifetime with cost constraints, FSDS and O3D attempt to

Table 2
Parameters of the simulated WSNs.

Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
τ_1	70%	L	512 (bits)	δ^2	10	r	100 (m)
Q_{SN}	1500	E^{SN}_{init}	3000 (J)	Q^{M}_{RN}	5	PDN	20%
ε_1	50×10^{-9} (J/bit)	V	300	k	3	Q_{RN}	15
ε_2	10×10^{-12} (J/bit/m^2)	P_{min}	−104 (dB)	g^{SN}_i	10 (byte/round)	C^{SN}_i	1000 (byte/h)
β	50×10^{-9} (J/bit)	t_{round}	24 (h)	g^{RN}_i	100 (byte/round)	C^{RN}_i	2000 (byte/h)
n	4.8	PNF	20%	τ_2	70%	E^{RN}_{init}	3000 (J)

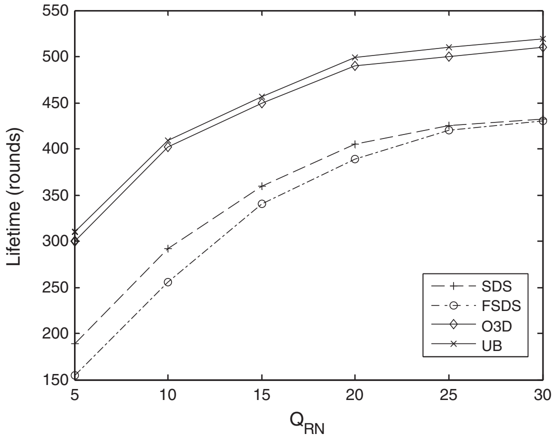

Fig. 1. Lifetime as a function of the number of RNs.

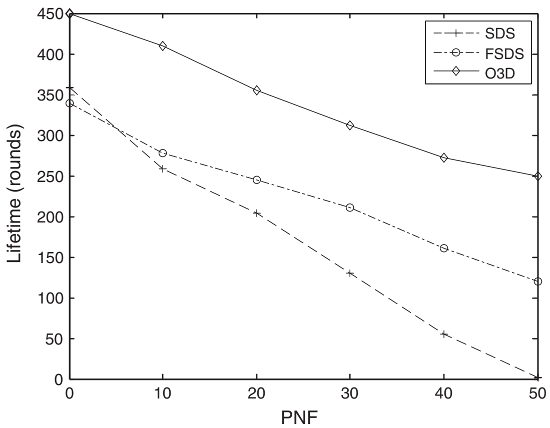

Fig. 2. Lifetime as a function of Probability of Node Failure (PNF).

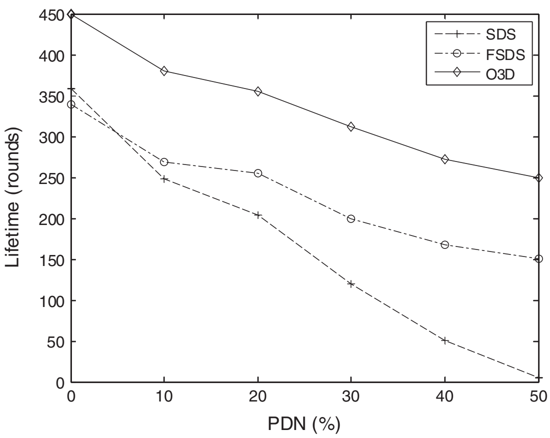

Fig. 3. Lifetime as a function of Probability of Disconnected Nodes (PDN).

find locations of these nodes such that the network lifetime is maximized under certain cost and fault-tolerance constraints. It is expected that the system lifetime (measured in rounds) improves as Q_{SN} and Q_{RN} increased in a given terrain. This behavior is depicted in Fig. 1 below. The more active nodes available in a given terrain the better the connectivity is and the lesser partitions are formed in the network. Consequently, this prolongs the network lifetime as mentioned earlier. The major effect of fault-tolerance constraints obviously appears in practical situations, where PNF and PDN are nonzero as depicted in Figs. 2 and 3. According to results obtained by applying Theorem 2, the mobility factor has a great influence on the efficiency of the proposed two-phase deployment scheme by which we can achieve an average of 97% of the UB. The small difference between results obtained by O3D and UB is due to the assumed probabilistic communication model that causes additional transmitting and receiving processes when bad channel conditions are experienced. Moreover, comparing the number of rounds achieved by the SDS and the FSDS strategies with those achieved by the UB, clearly shows the performance gains achieved by the O3D scheme over the two.

It is shown in Fig. 1 that the SDS and FSDS schemes were only able to achieve an average of 50% of the UB val-

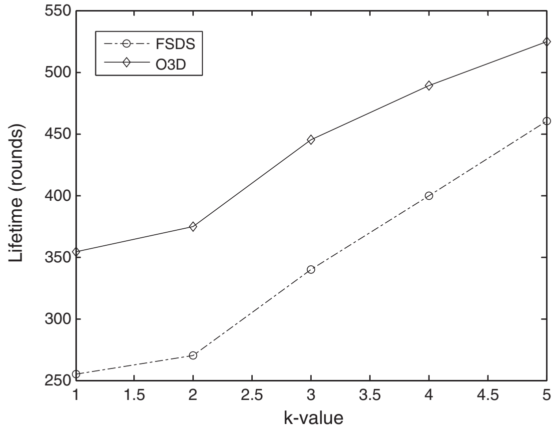

Fig. 4. Lifetime as a function of fault-tolerance level k.

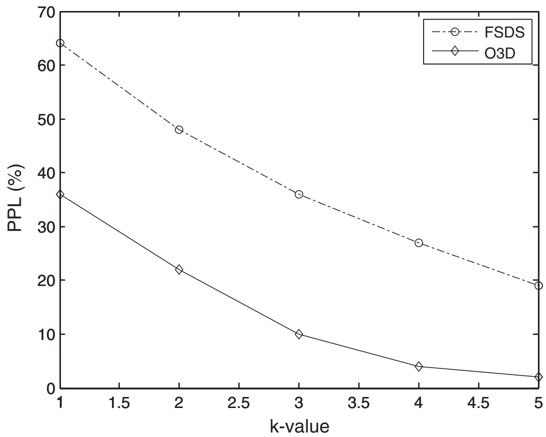

Fig. 5. Packet loss as a function of the fault-tolerance level k.

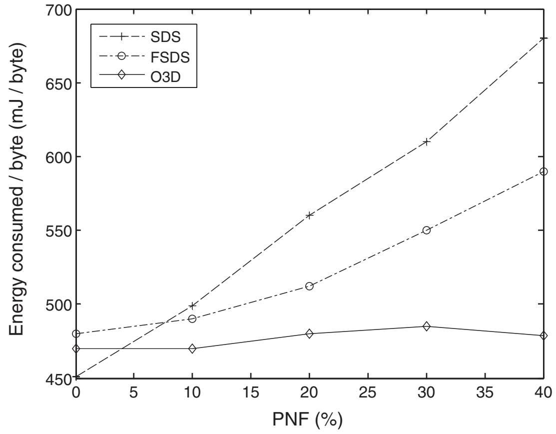

Fig. 6. Energy consumed vs. the Probability of Node Failures.

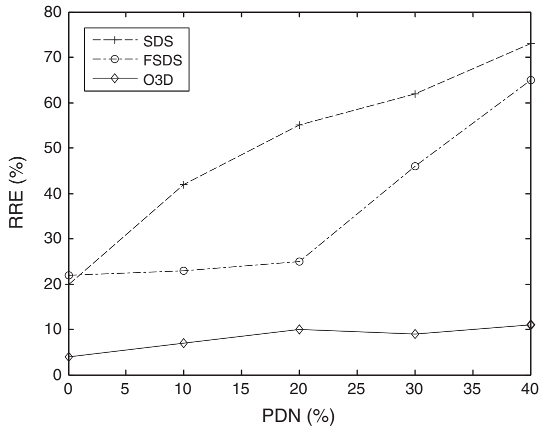

Fig. 7. Ratio of Remaining Energy (RRE) vs. PDN.

ues. Fig. 2 compares the PNF with respect to average lifetime measured in rounds assuming $Q_{RN} = 15$ and $Q_{SN} = 100 \times Q_{RN}$. Apparently, FSDS achieves better results than SDS. This is also the case when PDN is compared against average lifetime in Fig. 3. Fig. 4 compares the FSDS and the O3D in terms of the network lifetime under different fault tolerance levels. PDN and PNF are set to 20%. It can be seen that larger k values, about 80% more, are needed for the FSDS to give the same lifetime as the O3D. This feature is appealing in situations where node mobility is either too expensive or even impossible. A similar conclusion can be drawn from Fig. 5 for the PPL performance. Fig. 6 shows the monotonically increasing relation between the average energy consumed and the PNF. It can be seen that the higher the percentage of faulty nodes the more energy is consumed. This causes poor connectivity as well as extended communication delays, which are undesirable properties in OEM applications. However, combining the mobility feature with the fault-tolerance constraint, in O3D, leads to stable energy consumption per byte under varying PNF values. Fig. 7 illustrates the dependence between the RRE and the PDN. It shows that the higher the PDN, the energy is left unused, i.e. remained in the nodes while the network is deemed dead. This causes degraded connectivity which leads to more partitions, thus terminating the network's lifetime. Fig. 7 also shows how the RRE remains stable as long as the network has enough fault-tolerance level. For example, the RRE of the networks generated by FSDS remains around 23% as long as the PDN is less than or equal to 20%. For PDN > 20%, the RRE rapidly increases as the PDN increase. However, this problem does not appear with the O3D strategy due to its ability to replace the deployed nodes at the beginning of each round. Choosing an appropriate k-value highly dependents on the PNF and PDN as illustrated in Fig. 8. For low PNF, a low redundancy level is needed. On the other hand, k must be at least equal to 4 to guarantee the network functionality for at least 300 rounds in environments with a 50% PNF. Similar results can be obtained to determine the choice of k under specific PDN values.

5. Conclusions

This paper proposed a jointly energy-efficient and k-fault-tolerant node deployment strategy for heterogeneous WSNs. Intensive simulations showed that jointly considering energy-efficiency and fault-tolerance in node deployment can dramatically increase the network lifetime in OEM applications. To maintain these two objectives during the operation time, a certain number of MRNs is used. These MRNs can be relocated based on decentralized deci-

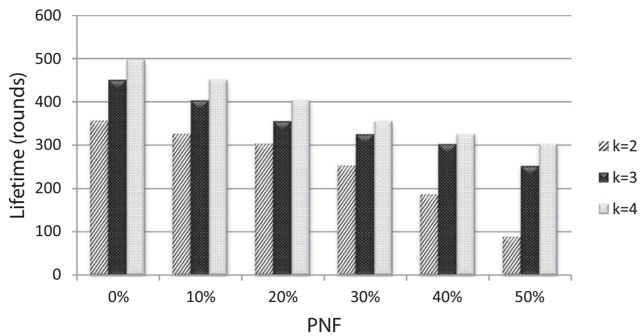

Fig. 8. Lifetime vs. the probability of node failure under three different fault-tolerance levels; $k = 2$, 3, and 4.

sions made at certain points in time such that energy efficiency and fault tolerance are maintained. To find the optimal positions of these MRNs, an optimization problem is formulated where the objective is to jointly minimize the total energy consumed and maximizing the minimum residual energy. It was shown that the proposed combined scheme with fault tolerance, energy efficiency, and MRNs, outperforms the previously proposed schemes in withstanding the harsh OEM conditions causing both, node and link failures. We evaluated the proposed scheme extensively based on real outdoor dataset, and presented a distributed implementation of WSN in OEM. This implementation introduces efficient load-balanced reporting, and shows that an improvement of 40% in the network lifetime can be achieved in comparison to classical approaches in the literature, where the three factors: *fault tolerance*, *energy efficiency*, and *MRNs* are ignored.

References

[1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, A survey on sensor networks, IEEE Commun. Mag. 40 (8) (2002) 102–114.
[2] M. Younis, K. Akkaya, Strategies and techniques for node placement in wireless sensor networks: a survey, Ad Hoc Netw. J. 6 (4) (2008) 621–655.
[3] F. Al-Turjman, H. Hassanein, M. Ibnkahla, Quantifying connectivity in wireless sensor networks with grid-based deployments, J. Netw. Comput. Appl. 36 (1) (2013) 368–377.
[4] P. Wang, Z. Sun, M. Vuran, M. Al-Rodhaan, A. Al-Dhelaan, I. Akyildiz, On network connectivity of wireless sensor networks for sandstorm monitoring, Comput. Netw. J. 55 (5) (2011) 1150–1157.
[5] I. Akyildiz, E. Stuntebeck, Wireless underground sensor networks: research challenges, Ad Hoc Netw. J. 4 (2006) 669–686.
[6] I. Akyildiz, D. Pompili, T. Melodia, Underwater acoustic sensor networks: research challenges, Ad Hoc Netw. J. 3 (2005) 257–279.
[7] M. Hashim, S. Stavrou, Measurements and modeling of wind influence on radiowave propagation through vegetation", IEEE Trans. Wireless Commun. J. 5 (5) (2006) 1055–1064.
[8] K. Akkaya, F. Senel, Detecting and connecting disjoint sub-networks in wireless sensor and actor networks, Ad Hoc Netw. J. 7 (7) (2009) 1330–1346.
[9] I. Stojmenovic, D. Simplot-Ryl, A. Nayak, Toward scalable cut vertex and link detection with applications in wireless ad hoc networks, IEEE Netw. 25 (1) (2011) 44–48.
[10] F. Wang, M. Thai, D. Du, On the construction of 2-connected virtual backbone in wireless networks, IEEE Trans. Wireless Commun. J. 8 (3) (2009) 1230–1237.
[11] H. Tan, I. Korpeoglu, I. Stojmenovic, Computing localized power efficient data aggregation trees for sensor networks, IEEE Trans. Parallel Distrib. Syst. 22 (3) (2011) 489–500.
[12] D. Pompili, T. Melodia, I. Akyildiz, Three-dimensional and two-dimensional deployment analysis for underwater acoustic sensor networks, Ad Hoc Netw. J. 7 (4) (2009) 778–790.
[13] G. Tolle, J. Polastre, R. Szewczyk, D. Culler, A microscope in the redwoods, in: Proc. ACM Conf. on Embedded Networked Sensor, San Diego, USA, 2005, pp. 51–63.
[14] B. Son, Y. Her, J. Kim, A design and implementation of forest-fires surveillance system based on wireless sensor networks for South Korea mountains, Int. J. Comput. Sci. Netw. Secur. 6 (9) (2006) 124–130.
[15] F. Al-Turjman, H. Hassanein, M. Ibnkahla, Connectivity optimization with realistic lifetime constraints for node placement in environmental monitoring, in: Proc. IEEE Conference on Local Computer Networks (LCN), Zürich, Switzerland, 2009, pp. 617–624.
[16] A. Erman, L. Hoesel, P. Havinga, J. Wu, Enabling mobility in heterogeneous wireless sensor networks cooperating with UAVs for mission-critical management, IEEE Trans. Wireless Commun. J. 15 (6) (2008) 38–46.
[17] P. Bellavista, A. Corradi, C. Giannelli, Mobility-aware middleware for self-organizing heterogeneous networks with multihop multipath connectivity, IEEE Trans. Wireless Commun. J. 15 (6) (2009) 22–30.
[18] K. Akkaya, M. Younis, A survey on routing protocols for wireless sensor networks, J. Ad Hoc Netw. 3 (3) (2005) 325–349.
[19] F. Al-Turjman, H. Hassanein, W. Alsalih, M. Ibnkahla, Optimized relay placement for wireless sensor networks federation in environmental applications, Wireless Commun. Mobile Comput. 11 (12) (2011) 1677–1688.
[20] K. Xu, H. Hassanein, G. Takahara, Q. Wang, Relay node deployment strategies in heterogeneous wireless sensor networks, IEEE Trans. Mobile Comput. 9 (2) (2010) 145–159.
[21] Q. Wang, K. Xu, G. Takahara, H. Hassanein, Device placement for heterogeneous wireless sensor networks: minimum cost with lifetime constraints, IEEE Trans. Wireless Commun. J. 6 (7) (2007) 2444–2453.
[22] W. Alsalih, H. Hassanein, S. Akl, Routing to a mobile data collector on a predefined trajectory, in: Proc. IEEE International Conference on Communications (ICC), Dresden, Germany, 2009, pp. 1–5.
[23] M. Dawande, R. Prakash, S. Venkatesan S. Gandham, Energy efficient schemes for wireless sensor networks with multiple mobile base stations, in: Proc. IEEE Global Telecommunications Conference (GLOBECOM), San Francisco, US, 2003, pp. 377–381.
[24] A. Azad, A. Chockalingam, Mobile base stations placement and energy aware routing in wireless sensor networks, in: Proc. IEEE Wireless Communications and Networking Conference (WCNC), Las Vegas, NV, 2006, pp. 264–269.
[25] M. Ishizuka, M. Aida, Performance study of node placement in sensor networks, in: Proc. International Conference on Distributed Computing Systems Workshops (ICDCS), Tokyo, Japan, 2004, pp. 598–603.
[26] B. Hao, H. Tang, G. Xue, Fault-tolerant relay node placement in wireless sensor networks: formulation and approximation, in: Proc. Workshop on High Performance Switching and Routing (HPSR), Phoenix, USA, 2004, pp. 246–250.
[27] K. Akkaya, M. Younis, M. Bangad, Sink repositioning for enhanced performance in wireless sensor networks, Comput. Netw. 49 (4) (2005) 512–534.
[28] F. Al-Turjman, H. Hassanein, M. Ibnkahla, Efficient deployment of wireless sensor networks targeting environment monitoring applications, Comput. Commun. J. 36 (2) (2013) 135–148.
[29] T. Rappaport, Wireless Communications: Principles and Practice, second ed., Prentice Hall, Upper Saddle River, USA, 2002.
[30] J. Rodrigues, S. Fraiha, H. Gomes, G. Cavalcante, A. De Freitas, G. De Carvalho, Channel propagation model for mobile network project in densely arborous environments, J. Microwaves Optoelectron. 6 (1) (2007) 189–206.
[31] H.O. Tan, I. Körpeoglu, Power efficient data gathering and aggregation in wireless sensor networks, ACM SIGMOD Rec. 32 (4) (2003) 66–71.
[32] H.O. Tan, I. Korpeoglu, I. Stojmenovic, Computing localized power-efficient data aggregation trees for sensor networks, IEEE Trans. Parallel Distrib. Syst. 22 (3) (2011) 489–500.
[33] S. Hussain, O. Islam, An energy efficient spanning tree based multi-hop routing in wireless sensor networks, in: Wireless Communications and Networking Conference (WCNC), IEEE, 2007, pp. 4383–4388.
[34] K. Kalpakis, K. Dasgupta, P. Namjoshi, Maximum lifetime data gathering and aggregation in wireless sensor networks, in: Proceedings of the 2002 IEEE International Conference on Networking (ICN'02), August 2002, pp. 685–696.

Fadi Al-Turjman is an assistant professor at the School of Engineering, University of Guelph, Canada. He is working in the area of wireless networks architectures, deployments, and performance evaluation. He obtained his Ph.D. in Computer Science from Queen's University in 2011. He received his B.Sc. (honors) and M.Sc. (honors) degrees in computer engineering from Kuwait University in 2004 and 2007, respectively. From 2005 to 2007, he was a researcher and teacher at the departments of information science and computer engineering in Kuwait University. During this period, he intensively worked on developing digital circuits and wireless sensor nodes. Since 2011, he is a research and teaching associate at Queen's University. He has authored and/or co-authored more than 40 reputable journal and international conference papers, in addition to chairing a number of workshops in international symposia and conferences; including the FTRA IET in MUSIC 2012, the IEEE WLN in LCN 2012 and 2013, and the IEEE G-IoT in GLOBECOM 2012.

Hossam S. Hassanein is a leading research in the areas of wireless and mobile networks architecture, protocols and services. His record spans more than 500 publications in journals, conferences and book chapters, in addition to numerous keynotes and plenary talks in flagship venues. He has received several recognition and best papers awards at top international conferences. He is also the founder and director of the Telecommunications Research Lab at Queen's University School of Computing, with extensive international academic and industrial collaborations. He is a senior member of the IEEE, and is a former chair of the IEEE Communication Society Technical Committee on Ad-hoc and Sensor Networks (TC AHSN). He is an IEEE Communications Society Distinguished Speaker (Distinguished Lecturer 2008–2010).

Mohamad Ibnkahla received his Ph.D. degree from the Institut National Polytechnique of Toulouse, France in 1996. He joined Queen's University, Canada, in 2000, where he is an Associate Professor. He led several projects applying Wireless Sensor Networks in various areas such as environment monitoring, wildlife tracking, pollution detection and control, food traceability and safety risk monitoring, highway safety, intelligent transportation and water management. He has published Signal Processing for Mobile Communications Handbook, CRC Press, 2004, Adaptive Signal Processing in Wireless Communications, CRC Press, 2008, Adaptation and Cross-layer Design in Wireless Networks, CRC Press, 2008, and Cognitive Wireless Sensor Networks (2011).