Computer Communications 36 (2013) 135–148

Contents lists available at SciVerse ScienceDirect

Computer Communications

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

Efficient deployment of wireless sensor networks targeting environment monitoring applications

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

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

ARTICLE INFO

Article history:
Received 8 February 2011
Received in revised form 23 June 2011
Accepted 29 August 2012
Available online 21 September 2012

Keywords:
Connectivity
Lifetime
Deployment
Relay nodes
Wireless sensor networks

ABSTRACT

Maximizing network connectivity while maintaining a useful lifetime period without exceeding cost constraints is a challenging design objective for wireless sensor networks. Satisfying such objective becomes even a more intricate task with 3-D setups and harsh operational conditions found in typical large scale environment monitoring applications. While much work has been performed in environment monitoring, only few have addressed the unique characteristics of such applications. In this paper, we introduce a novel 3-D deployment strategy, called Optimized 3-D deployment with Lifetime Constraint (O3DwLC), for relay nodes in environmental applications. The strategy optimizes network connectivity, while guarantying specific network lifetime and limited cost. Key to our contribution is a very limited search space for the optimization problem, in addition to a revised definition for network lifetime that is more appropriate in environment monitoring. The effectiveness of our strategy is validated through extensive simulations and comparisons, assuming practical considerations of signal propagation and connectivity.

1. Introduction

Wireless sensor networks (WSNs) enable long-term environment monitoring at scales and resolutions that are difficult, if not impossible, to obtain using conventional techniques. WSNs can be re-tasked after deployment in the field based on changes in the environment, conditions of the sensor network itself, or scientific endeavor requirements [3]. However, most environmental applications require all seasons data gathering under inconsiderate conditions, and thus long-lasting tightly-connected networks in the monitored field are mostly wanted in order to have better understanding of the monitored phenomena such as life cycle of huge redwood trees [34], or to satisfy specific application objectives for several years as in forestry fire detection [31].

One of the most efficient techniques used for maintaining long-lasting connected networks is achieved by deploying Relay Nodes (RNs) [19]. Relay nodes can have extra storage space and much more powerful transceivers in order to forward sensed data for long distances in huge monitored sites, and thus energy at Sensor Nodes (SNs) is saved for further data sensing and gathering. Nevertheless, deployment of relay nodes [11,24] in environmental applications is a challenging problem due to harsh environments and required 3-D setups.

Harshness of the environment arises because of the nature of outdoor monitoring applications where sensor networks may work under heavy rain, extreme temperature variations and sometimes during stormy days destroying the deployed nodes and/or their communication links (edges). Nodes and communication links may also be destroyed by unexpected visitors such as birds and wild fauna. Moreover, due to dense trees and growing foliage, communication links are attenuated and connectivity availability is affected. Hence, nodes and links are prone to several risks leading to high probabilities of failures and many nodes may become disconnected which also degrades the overall network lifetime. As a reason of that, we are characterizing harshness of the monitored environment by the Probability of Node Failure (PNF) and Probability of Disconnected Nodes (PDN). Due to high PNF and PDN in environmental applications, it is reasonable to have redundant nodes in the deployed sensor networks [23]. In the scope of this paper, a *redundant* node is the node which can be removed from the network without affecting the targeted data. Contrarily, *irredundant* node is defined as a unique source of information in the monitored site that cannot be recovered by other nodes in the network.

Meanwhile, deployment in environmental applications becomes more challenging when 3-D setups are required. In environmental applications, relay nodes are not only forwarding data from different variations in the horizontal plane, but also from different vertical levels (e.g. on trees, at soil surface and even underground). For instance, in monitoring the gigantic redwood trees in California, some experiments required sensor placements at different heights on these trees spanning a range of several tens of meters

* Corresponding author. Address: Department of Computer Science, King Saud University, Riyadh, Saudi Arabia.

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

0140-3664/$ - see front matter © 2012 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.comcom.2012.08.021

[34]. There is also increased interest in 3D environmental applications such as CO_2 flux monitoring and imagery [31], where sensors are placed at different vertical levels to fulfill coverage and data accuracy requirements. Accordingly, communication links between deployed sensor nodes has to be considered in 3-D space rather than 2-D plane only, raising more complexity in terms of the deployed network connectivity [12,26].

Nonetheless, existing deployment schemes for such applications [28,29,31] are not always based on sound connectivity models, but rather on simplified lifetime models. They have not efficiently addressed the problem of connectivity in 3-D space, which is a natural model in environmental applications. For example, deployment in Ref. [18] focused on connectivity in 2-D outdoor applications. Deployment algorithms are proposed in Refs. [25,33] to guarantee the 2-D connectivity and/or ensure survivability in case of node failure without lifetime considerations. In contrast, the deployment in Ref. [15] aimed at maximizing the network lifetime under specific energy budget. The energy provisioning and relay node placement are formulated as a mixed-integer linear programming problem in 2-D plane. Heuristic algorithms are then introduced to overcome the computational complexity. Nevertheless, relay nodes in Ref. [15] are assumed to adapt their transmission range to reach any other node in the network and hence, connectivity is not considered as an issue (which is not practical in environmental applications). A hybrid approach has been proposed in Ref. [35] to balance connectivity and lifetime in 2-D outdoor deployments. Even in the most recent papers in environmental WSNs applications, including volcano [30] and harsh industrial [21] environment monitoring, network connectivity is considered in a 2-D plane with the assumption of a very basic binary communication disc model which is not the case in reality. Thus, these approaches are very prone to failure in practical large-scale environmental applications.

It is worth mentioning that there are some attempts towards the 3-D deployment. As an example, authors in Ref. [18] studied the effects of sensing and communication range on connectivity in 3-D space. However, connectivity optimization with lifetime constraints has not been investigated so far in 3-D environmental applications. As noted in Ref. [37], many of the popular deployment strategies which are optimally solved in polynomial time in 2-D plane, become NP-Hard in 3-D settings.

We also remark that relay nodes in environment monitoring are generally more expensive than sensing nodes due to the cost of the wide-range transceivers used to cover large-scale areas, like forests and cities. Therefore, *efficient* 3-D deployment must maintain connectivity and lifetime that limits the number of these expensive nodes. In *efficient* environmental deployments it is also undesirable to re-visit the monitored sites (e.g. for node replacement or battery recharging). Therefore, the deployed wireless sensor network must be guaranteed to function for a pre-specified lifetime period. For more accurate and practical lifetime guarantees, environment-specific lifetime definition should be considered. Even though there are several definitions for WSNs lifetime in the literature, there is no agreement on a definition for lifetime in environmental applications. Inappropriate definition might lead to incorrect lifetime estimation and hence, may cause a waste of resources.

In this paper, we investigate an efficient way for the relays placement that address aforementioned challenges and desired WSNs features in environmental applications. Such nodes' placement problem has been shown in Ref. [5] to be NP-hard. Finding non-optimal approximate solutions is also NP-hard in some cases [13]. To address this complexity, we propose an efficient two-phase relay node deployment in 3-D space, called O3DwLC strategy. The first phase of O3DwLC is used to setup a connected network backbone using minimum number of relay nodes for cost efficiency. In the second phase, we aim at finding a set of a relatively small number of candidate positions, such that we optimize the relay nodes placement on these positions to achieve the maximum backbone connectivity for guaranteed lifetime period within a limited cost budget. This two-phase deployment scheme will provide a reliable interaction with the network end users monitoring outdoor environments. It will optimize the relay node deployment in forests to detect fires and report wild life activities, and in water-bodies to record events concerning floods, water pollution, coral reef conditions and oil spills, in addition to targeting other rural and hazardous areas such as deserts, polar and volcanic terrains, and battle fields.

Major contributions of this paper are listed as follows. We explore the most suitable lifetime definition in environment monitoring. The appropriateness of the proposed definition is evaluated and compared to other lifetime definitions in the literature based on harsh environmental characteristics. We introduce a generic 3-D relay node placement problem, which aims at maximizing connectivity with constraints on wireless sensor network cost and lifetime. We propose an efficient two-phase solution for the 3-D deployment problem, which considers a limited search space, generic communication model, most appropriate lifetime definition, and harsh operational conditions. Performance of the proposed two-phase solution is evaluated and compared to currently used strategies in environmental applications in the presence of varying probabilities of node failure and disconnectivity.

The remainder of this paper is organized as follows. In Section 2 related work is outlined. Practical system models and placement problem are presented in Section 3. In Section 4 our two-phase deployment strategy is described. The performance of the proposed strategy is evaluated and compared to other deployment strategies in Section 5. Finally, conclusions and future work are given in Section 6.

2. Related work

Extensive work has been reported in the literature relating to relay node deployment strategies which are classified into random vs. grid-based deployments [37]. In random deployment, nodes are randomly scattered and are organized in an ad hoc manner. While in grid-based deployment, nodes are placed on grid vertices leading to more accurate positioning and data measurements. In addition, applying grid deployments in 3-D space has several other benefits. Such deployments precisely limit the search space of relay nodes positions and possible paths between them, as well. Thus, the formulation of the placement optimization problem is simplified. Moreover, grid models can reflect channel conditions using specific signal propagation models that consider harsh environment characteristics, and hence better connectivity can be maintained. Due to the interest of environmental applications in the exact physical positioning of sensor and relay nodes, grid-based deployment is the most appropriate and is adopted in this work. However, more efficiency is required in grid-based deployments to enhance the deployed nodes connectivity and limit their huge number of candidate positions (or grid vertices) in large-scale environmental applications.

Connectivity of the deployed grid-based network could be presented as k-connectivity. k-Connectivity has two different meanings, namely, k-path connectivity and k-link connectivity [17]. The k-path connectivity means that there are k independent paths between every pair of nodes, while k-link connectivity means that each node is directly connected to k neighboring nodes. Nevertheless, a wireless sensor network could be disconnected even if k-link connectivity is satisfied. With k-path connectivity where $k \geqslant 1$, the network can tolerate some node and link failures. At the same time, the higher degree connectivity improves communication capacity among nodes. In some cases, it may not be necessary to maintain k-connectivity among all of the network nodes, but only among nodes which form the communication backbone of the network, and thus is adopted in this paper. Nodes constructing the network backbone are called *irredundant* (critical) nodes, in this research. Fail of connectivity between these nodes may lead to sever effects on the WSN performance such as network partitioning and data loss. Accordingly, node redundancy, in Ref. [8], is used to overcome such connectivity problems. Redundant nodes are deployed and the ones that are not being used for communication or sensing are turned off. When the network becomes disconnected, one or more of the redundant nodes is turned onto repair connectivity. In Ref. [22], the lowest number of redundant nodes is added to a disconnected static network, so that the network remains connected. Similarly, authors in [20] focus on designing an optimized approach for connecting disjointed WSN segments by populating the least number of relays. The deployment area is modeled as a grid with equal-sized cells. The optimization problem is then mapped to selecting the fewest count of cells to populate relay nodes such that all segments are connected. In addition, overlapping clusters of sensor nodes, which rely on the concept of redundancy as well, are used to enhance the network connectivity in [37]. In Ref. [1], a distributed recovery algorithm is developed to address 1- and 2-connectivity requirements. The idea is to identify the least set of nodes that should be repositioned in order to reestablish a particular level of connectivity. Nonetheless, redundant nodes deployment becomes an intricate task in huge 3-D spaces where numerous positioning options are possible with different connectivity levels (degrees). Therefore, more efforts are required to optimize the deployment process under such circumstances.

Meanwhile, grid-based deployment should not only guarantee connectivity in harsh environment monitoring but also should guarantee specific network lifetime. Lifetime has several definitions in the literature. One of the most common lifetime definitions states: "it is the time till the first node death occurs" [15]. Such definition may not be appropriate if we are monitoring forest temperature or humidity because if a node dies, we can still receive similar (or redundant) information from other nodes in the same area. Therefore, this definition may only be considered as a lower bound of other lifetime definitions and should not reflect the actual network lifetime. Similarly, the definition, which says that "it is the time till the last node death occurs", can serve as an upper bound of other definitions. Both of these definitions are unrealistic for environmental applications. Another way to define lifetime could rely on the percentage of alive nodes (which have enough energy to accomplish their assigned tasks) [15]. But choosing such percentage threshold is usually arbitrary and does not reflect the application requirements. In addition, being alive does not mean that the node is still connected. Therefore, other approaches to evaluate network lifetime are relying on connectivity and network partitions [9]. Still these definitions do not satisfy the environment monitoring applications. For instance, if a set of nodes are destroyed because of the movement of wild animals or falling trees in the monitored site, the network can still be considered functional as long as some other nodes are still alive and connected. In Refs. [6,32], the lifetime definitions rely on the percentage of covered area in the monitored site. But they are not suitable for environmental applications which are data-driven. In data-driven applications, we are more interested in sampling data from the monitored site rather than providing full coverage. Tightly connecting these data gathering nodes can reduce redundant information transmissions and hence, prolong the overall network lifetime.

Motivated by the benefits of device heterogeneity, as well as, the 3-D grid model our research provides an efficient grid-based deployment for provisioning WSNs of maximum backbone (critical nodes) connectivity degree under lifetime constraints and limited cost budget in environmental applications. For more efficiency and unlike other grid-based deployments, we are finding the most feasible grid vertices to be searched for the optimal deployment rather than searching a massive number of grid vertices in large-scale environmental applications.

3. System models and problem definition

In this section, we outline our assumed wireless sensor network models, in addition to introducing a general definition for the targeted relay node deployment problem. We assume hierarchical network architecture to address the node heterogeneity problem. A graph topology is considered for easy network extend, and accurate (mathematical) connectivity computation. Furthermore, a detailed discussion of the utilized cost and communication models is proposed, and the appropriateness of the considered lifetime definition is examined.

3.1. Network model and placement problem

In this paper, a two-layer hierarchical architecture is assumed as a natural choice in large-scale environmental applications, in addition to providing more energy-efficient deployment plan. The lower layer consists of sensor nodes that sense the targeted phenomena and send measured data to Cluster Heads (CHs) in the upper layer, as shown in Fig. 1(a). Usually these sensor nodes have fixed and limited transmission ranges and do not relay traffic in order to conserve more energy. The upper layer consists of cluster heads and relay nodes which have better transmission range (=r) and communicate periodically with the base station to deliver the measured data in the lower layer. Cluster heads aggregate the sensed data and coordinate the medium access, in addition to supporting relay nodes in relaying data from other CHs to the BS in the upper layer. Assuming sensor nodes have enough energy to perform their effortless tasks, we focus this work on the upper layer devices which are relay nodes and cluster heads. The topology of the upper layer is modeled as a graph $G = (V,E)$, where $V = \{n_0, n_1, \ldots, n_{nc}\}$ is the set of n_c candidate grid vertices, E is the set of edges in graph G, and $(i,j) \in E$, if nodes at n_i and n_j have enough probabilistic connectivity percentage to establish a communication link (edge). We remark that deployment of relay nodes in this research is independent of the underlying Medium Access Control (MAC) protocol, where we assume a transmission rate limit T for each node during a one time unit (measured in days). This limit can be adjusted to comply with any MAC protocol. For simplicity and without losing generality, we assume S-MAC protocol [35] is handling the medium access in this research with a 50% duty cycle and only six bytes controlling fields in the exchanged packets for more energy savings. Moreover, the assumed traffic generation model is flexible enough to support different requirements of numerous WSNs applications. As we have specific parameters; T and Y, controlling the assumed packet transmission rate and arrival rate, respectively. For example, if we increase T, packets will be generated more frequently. If we decrease Y, packets will spend more time to reach the destination, and thus more congested network can be experienced. In this research, we assume that the arrival rate Y is following a passion process which is very nature in WSNs simulations.

Fig. 1(b) depicts the 3-D grid model assumed in this paper, where the grid edge length is supposed to be equal to a relay node transmission range r. It is assumed that all relay nodes have a common transmission range r. We remark that our deployment planning is applicable for other types of grid models, not only the cubic one. In this cubic grid model, each Sensor Node (SN) is placed near to phenomena of interest for more accurate estimates in terms of the spatial properties of the collected data. Cluster Heads (CHs) are then placed on the most appropriate grid vertices; which

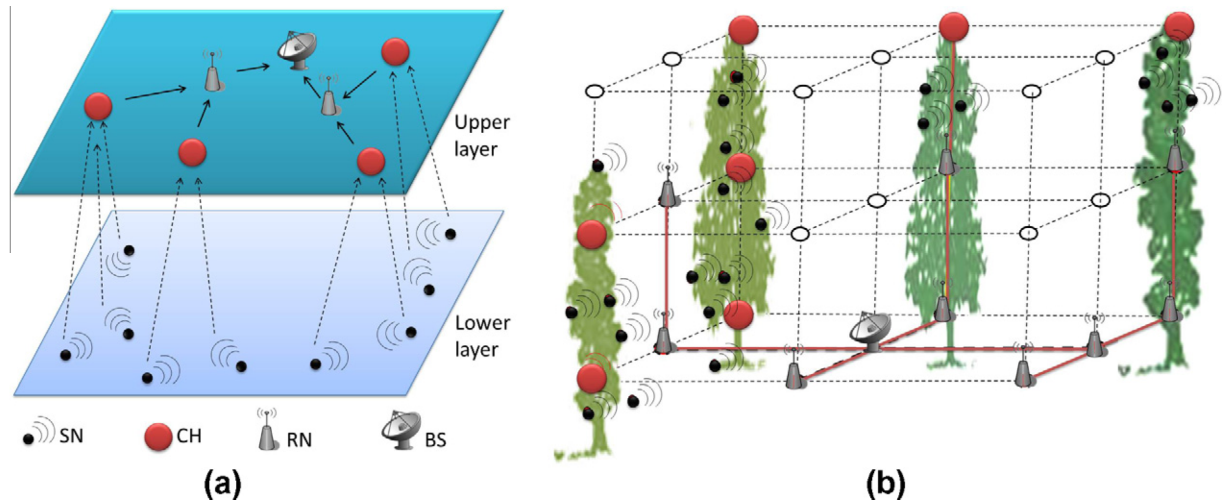


Fig. 1. (a) Two layer hierarchical architecture and (b) Cubic 3-D grid model for the targeted wireless sensor network deployment where dashed lines and empty circles represent grid edges and vertices, respectively.

can serve the largest number of sensor nodes distributed around each cluster head. The base station is placed based on the application requirements in a fixed position and it is the data sink for the system. Then, we seek to optimize the relay nodes positions on the 3-D grid to get cluster heads connected to the base station efficiently; in terms of cost and network lifetime. Hence, we define a general relay node placement problem in environmental wireless sensor networks as follows:

Problem statement definition: Given a specific sensing task with pre-specified SNs, CHs and BS locations, determine the positions of RNs so that connectivity between CHs and BS is maximized while lifetime and cost constraints are satisfied.

3.2. Cost and communication models

Device cost in environmental applications depends on its functionalities and hardware components. The more functionality the device has, the more complex and expensive it is. As relay nodes are assumed to have more functionality and dominate other devices in terms of transmission range, the cost is modeled in this paper by the number of relay nodes placed in the monitored site. We assume identical cost for these RNs.

For the communication model, we consider a probabilistic connectivity between the deployed devices, in which wireless signals not only decay with distance, but also are attenuated and reflected by surrounding obstacles including trees, animals, hills, etc. Accordingly, the communication range of each device must be represented by an arbitrary shape as depicted in Fig. 2. For realistic estimation of the ability to communicate within this arbitrary shape, we need a signal propagation model that reflects the effects of the surrounding obstacles and the environment characteristics on the propagated signals. This model can describe the path loss[1] in the monitored environment as follows [28].

$$P_r = K_0 - 10\gamma \log(d) - \mu d \qquad (1)$$

where P_r is the received signal power, d is the Euclidian distance between the transmitter and receiver, γ is the path loss exponent calculated based on experimental data, μ is a random variable that follows a log-normal distribution function with zero mean and variance δ^2 to describe signal attenuation effects[2] in the monitored

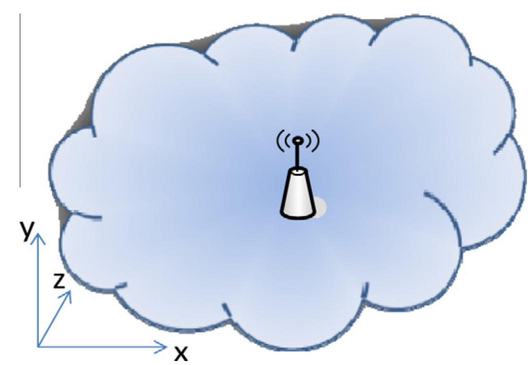


Fig. 2. An arbitrary shape of the communication range in 3-D space, due to attenuation and shadowing affecting outdoor wireless signals.

site, and K_0 is a constant calculated based on the transmitter, receiver and monitored site mean heights.

Let P_r equal the minimal acceptable signal level to maintain connectivity. Assume γ and K_0 in Eq. (1) are also known for the specific site to be monitored. Thus, a probabilistic communication model which gives the probability that two devices separated by distance d can communicate with each other is given by

$$P_c = Ke^{-\mu d^{\gamma}} \qquad (2)$$

where $K_0 = 10\log(K)$.

Thus, the probabilistic connectivity P_c is not only a function of the distance separating the wireless nodes but also a function of the surrounding obstacles and terrain, which can cause shadowing and multipath effects (represented by μ). Thus, the ability to communicate between two nodes is defined as follows:

Probabilistic connectivity definition: Two nodes (devices) i and j, separated by distance d, are connected with a threshold parameter $\tau (0 \leqslant \tau \leqslant 1)$, if $P_c(i,j) \geqslant \tau$.

Note that this communication model is generic in terms of the parameters (K, μ, τ, and γ), which specify the surrounding environment characteristics. Setting these parameters to values obtained from experimental data would provide more practical connectivity estimation and thus, more efficient deployment planning.

3.3. Lifetime model

Models in the literature differ in the way they consider a wireless sensor network to be still operational. These models can rely

[1] Path loss is the difference between the transmitted and received power of the signal.

[2] Wireless signals are attenuated because of shadowing and multipath effects. This refers to the fluctuation of the average received power.

on connectivity of the deployed nodes or on percentage of alive nodes (which have enough energy to accomplish their assigned tasks) in the network. *Connectivity-based* (*CB*) and *Percentage of Alive Nodes (PAN)* models are defined as follows, respectively:

Connectivity-based definition: *Lifetime of a WSN is the time span from deployment to the instant when a network partition occurs.*[3]

Percentage of alive nodes definition: *Lifetime of a WSN is the time span from deployment to the instant when the percentage of alive nodes falls below a specific threshold.*

However, losing a few nodes may not significantly affect the overall wireless sensor network performance especially when redundant nodes and communication links (edges) are used in tolerating high probabilities of disconnected nodes, as they did in environmental applications. Generally, in environment monitoring, several nodes are assigned to measure single specific criteria of the monitored space, such as temperature in forestry fire detection [31]. Consequently, the concept of node redundancy should be addressed. In addition, lifetime models relying on the aforementioned definitions do not take into consideration the node type which could be cluster head, relay node or sensor node. Therefore, we propose the following *Environment-specific (Env.)* lifetime definition.

Environment-specific lifetime definition: *Lifetime of a WSN is the time span from deployment to the instant when the percentage of alive and connected irredundant nodes is below a pre-defined specific threshold.*

Using this definition we benefit from device redundancy by considering the network to be operational as long as a specific percentage of cluster heads providing the targeted data, are still alive. These cluster heads need not only be alive, but also must be connected to the base station via single or multi-hop path(s). Note that a cluster head i is connected to another node j, if $P_c(i,j) \geqslant \tau$, according to the *probabilistic connectivity* definition.

In order to mathematically translate the aforementioned lifetime definitions, we assume *number of rounds* for which a wireless sensor network can stay operational as the unit measure of the network lifetime. A complete *round* is defined in this paper as the time span t_{round} in which each irredundant cluster head (i.e. responsible for different sensor nodes) transmits at least once to the base station without violating cutoff criteria of the lifetime definitions. t_{round} is identical for all rounds due to a constant data delivery assumed per round. In addition, we adopt the general energy consumption model proposed in [35], in which energy consumed for receiving a packet of length L is:

$$J_{rx} = L\beta \tag{3}$$

and the energy consumed for transmitting a packet of length L for distance d is:

$$J_{tx} = L(\varepsilon_1 + \varepsilon_2 d^{\gamma}) \tag{4}$$

where ε_1, ε_2 and β are hardware specific parameters of the utilized transceivers, and γ is the path loss exponent.

Based on Eqs. (3) and (4), in addition to knowing the initial energy E_i of each node with its relative position to other nodes, we can calculate the remaining energy E_r per node after the completion of each round by

$$E_r = E_i - TJ_{tx} - RJ_{rx} - AJ_a \tag{5}$$

where T, R and A are the arrival rates of transmitted, received and aggregated packets per round, respectively, that follows a Poisson distribution, and J_a is the energy consumed for a single packet aggregation. Considering E_r calculated in Eq. (5) and assuming the cutoff criterion associated with each lifetime definition is represented by a binary variable[4] C, we can calculate the total number of rounds for which a wireless sensor network can stay operational for.

To assess the *environment-specific* definition, we use simulation to compare it to *Connectivity-based* and *Percentage of alive nodes* definitions using four main performance metrics: (1) *Percentage of alive CHs,* (2) *Percentage of disconnected CHs/RNs,* (3) *Ratio of Remaining Energy (RRE),* and (4) *Total rounds*, which are described as follows. *Percentage of alive CHs* is the percentage of cluster heads which have enough energy to aggregate and forward data to the base station at least once. *Percentage of disconnected CHs/RNs* is the percentage of cluster heads and relay nodes which have enough energy to aggregate and forward data at least once but are not able to communicate with the base station. *Ratio of remaining energy* is the ratio of total energy amount still available at all nodes (CHs/RNs) to the total energy at deployment when the network is not operational. The network is not operational when the cutoff criterion of the lifetime definition is satisfied. Finally, *total rounds* is the total number of rounds in which a wireless sensor network can be considered operational. These four performance metrics are chosen to reflect the ability of *environment-specific* lifetime definition to: (1) accurately estimate the network lifetime and (2) effectively utilizing the network (i.e. maximize network operational time by delaying the assumption of the network death). Thus much better energy and resource utilization can be achieved.

Using Matlab, we simulate randomly generated wireless sensor networks which have the hierarchical architecture and the graph topology proposed in Section 3.A. Each generated network consists of 12 CHs and a total of 50–80 RNs which are randomly deployed on grid vertices in $700 \times 700 \times 200$ (m^3) 3-D space using Linear Congruential random number generator. The parameters used in the simulations are listed in Table 1.

Therein, τ is set to high value for practicality in simulating fluctuated and attenuated signals in environment monitoring applications [28]. For simplicity, we apply cubic 3-D grid model with identical grid edges of the length equal to 100 (m). We assume a predefined fixed time schedule for traffic generation (=100 packets per round from each CH) and a Probability of Node Failure (PNF) varying from 10–60%. We define the PNF as the probability of physical damage for each node in the network, which is very common in outdoor environment monitoring. Thus, a higher PNF indicates a higher possibility for the node to be damaged; while still having enough energy to sense and communicate. We intended to assume a very high PNF (up to 60%) to reflect some actual situations in outdoor environmental applications. Three different cutoff criteria are used for the simulated network to be considered operational. According to *Environmental-specific* (*Env.*) lifetime definition and the proposed network model, the network is still operational as long as the percentage of connected irredundant cluster heads which have enough energy to communicate with the base station is greater than or equal to 50%. The *CB* definition considers the network non operational when one or more irredundant cluster heads are unable to reach the base station. Finally, the network is not operational, based on *PAN* definition, when 50% or more of the nodes run out of energy. After the network is considered not operational, we measure the aforementioned performance metrics. This experiment is repeated 500 times. The average results are reported in Figs. 3–6. We remark that the average results hold a confidence interval no more than 5% of the average (over 500 runs) at a 95% confidence level.

Fig. 3 shows the percentage of disconnected nodes obtained when the network becomes non-operational. Obviously, *CB* definition underestimates the network lifetime by considering it non operational while it has very low percentage of disconnected nodes. Dependency of the disconnected nodes percentage on the

[3] Network partition occurs when one or more nodes are not able to communicate with the base station.

[4] The cutoff criterion is not satisfied and the network is still considered operational if $C = 0$ and vice versa if $C = 1$.

Table 1
Parameters of the simulated WSNs.

Parameter	Value	Parameter	Value
τ	70%	L	512 (bits)
n_c	110 (vertex)	E_i	15.4 (J)
ε_1	50e-9 (J/bit)	T	100 (packet/round)
ε_2	10e-12 (J/bit/m^2)	P_r	−104(dB)
β	50e-9 (J/bit)	t_{round}	24 (hour)
γ	4.8	K_0	42.152
δ^2	10	r	100 (m)
J_a	50e−7 (J)	PNF	10–60%

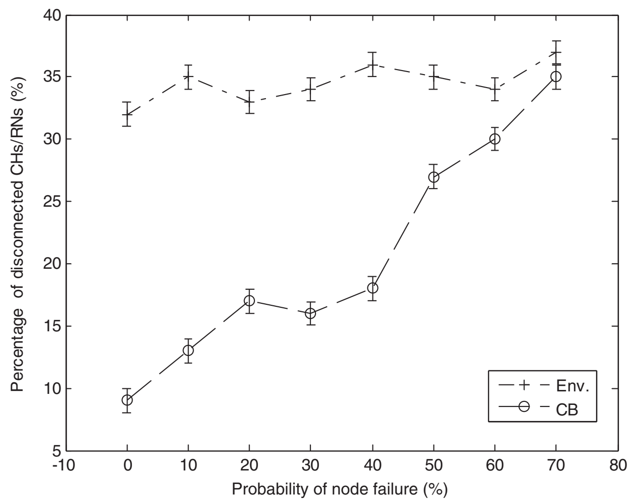


Fig. 3. Percentage of disconnected CHs/RNs vs. different probabilities of CH/RN failure.

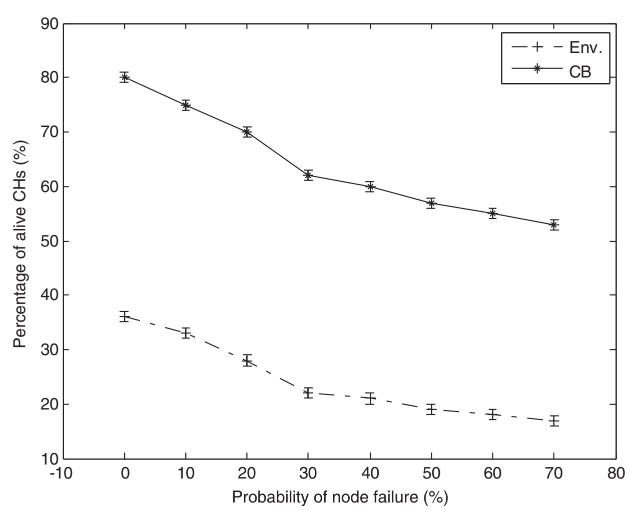


Fig. 4. Percentage of alive CHs vs. different probabilities of CHs/RNs failure.

PNF values, confirms its unsuitability for environmental applications. Using *Env.* definition we observe that the simulated wireless sensor networks can remain operational even when the percentage of disconnected nodes is much higher than the percentage achieved by the *CB* definition. Unlike *CB* definition, *Env.* definition shows very close percentages of disconnected nodes under different probabilities of node failure when the network is considered non operational. This steady state in lifetime estimation is preferred under the varying probabilities of node of failures in outdoor environments. In Fig. 4, percentage of alive CHs when the network

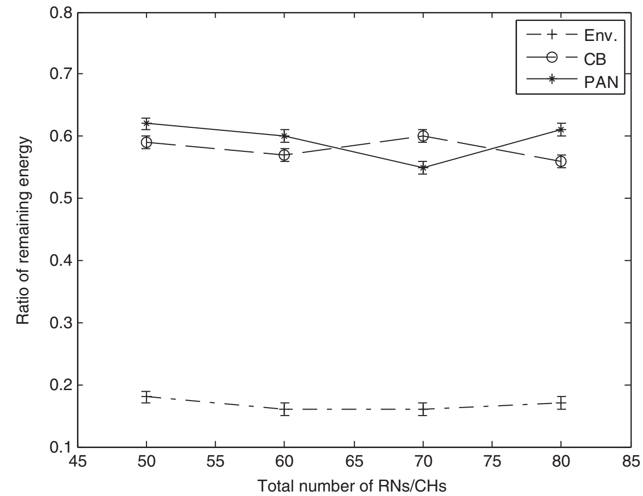


Fig. 5. Comparison of the three different definitions in terms of Ratio of Remaining Energy (RRE).

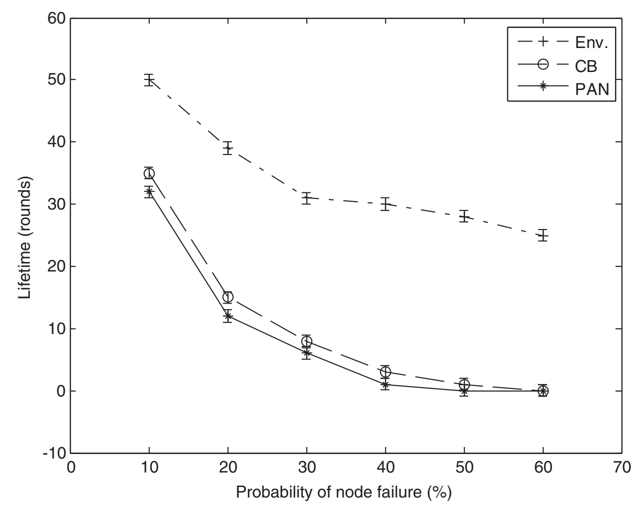


Fig. 6. Comparison of the three different definitions in terms of total rounds.

becomes non operational is shown. In this figure we can see how definitions relying on percentage of alive nodes waste network resources by stopping the network while it still has significant percentage of alive CHs. Furthermore, *PAN* definition considers that the network is not operational based on unpractical percentage of alive nodes which ignores node type (whether it is CH or RN) and does not differentiate between redundant and irredundant CHs. Therefore, *PAN* is not suitable for environmental applications. Fig. 5 shows that based on the *Env.* definition the wireless sensor network can remain operational even when the RRE is less than 20%. On the other hand, using *CB* and *PAN* definitions, the network is non operational even when it has around 60% of the initial deployed energy. Thus, Figs. 4 and 5 indicate how irrelevant lifetime definitions may lead to sever waste in resources (e.g. functional nodes, remaining unexploited energy, etc.) by assuming the network is dead while it stile have the ability to continue its assigned task. Fig. 6 depicts the total rounds counted based on the three lifetime definitions. It elaborates on how lifetime models relying on *CB* and *PAN* definitions can underestimate the overall network lifetime. It indicates the appropriateness of *Env.* definition in practice under harsh outdoor operational conditions. Consequently, *Env.* lifetime definition is imposed in our deployment strategy that is described in the following section.

4. Deployment strategy

The relay node placement problem proposed in this paper has infinitely large search space and finding the optimal solution is highly non-trivial. Therefore, we propose a 3-D grid model that limits the search space to a more manageable size. Grid models have well-organized vertices, distributed in regular lattice structures. These vertices can be organized in different structures (e.g. cubes, octahedrons, pyramids, etc.) in 3D space to provide more accurate estimates in terms of the spatial properties of the targeted data.

We assume knowledge of the 3-D terrain of the monitored site ahead of the deployment planning time. Hence, practical candidate positions on the grid vertices are pre-determined; non-feasible positions are excluded from the search space. We use candidate grid vertices to apply the deployment strategy in two phases. The first phase is used to place a minimum number of relay nodes on the grid vertices to establish a connected network. The second phase is used to choose the optimal positions of extra relay nodes required to maximize the network connectivity with constraints on cost and lifetime. This two-phase deployment strategy is called *Optimized 3-D Grid Deployment with Lifetime Constraint (O3DwLC)*.

4.1. First phase of the O3DwLC strategy

The first phase is achieved by constructing a connected Backbone (B) using First Phase Relay Nodes (FPRNs). Locations of these nodes are optimized in order to use minimum number of FPRNs that can connect cluster heads to the base station. Towards this end, we apply the Minimum Spanning Tree (MST) algorithm as described in Algorithm 1.

Algorithm 1. MST to construct the connected backbone B

1. **Function ConstructB** (IS: Initial Set of nodes to construct B)
2. **Input:**
3. A set *IS* of the CHs and BS nodes' coordinates
4. **Output:**
5. A set *CC* of the CHs, minimum RNs, and BS coordinates forming the network Backbone
6. **begin**
7. CC = set of closest two nodes in IS;
8. $CC = CC \cup$ minimum RNs needed to connect them on the 3-D grid;
9. $IS = IS - CC$;
10. N_d = number of remaining *IS* nodes which are not in *CC*;
11. $i = 0$;
12. **foreach** remaining node n_i in *IS* **do**
13. Calculate M_i: Coordinates of minimum number of RNs required to connect n_i with the closest node in *CC*.[a]
14. $i = i + 1$;
15. **end**
16. $M = \{M_i\}$
17. **while** $N_d > 0$ **do**
18. SM = Smallest M_i;
19. $CC = CC \cup SM \cup n_i$;
20. $IS = IS - n_i$;
21. $M = M - M_i$;
22. $N_d = N_d - 1$;
23. **end**
24. **end**

[a] This is achieved by counting the minimum number of adjacent grid vertices, which establish a path from the separated CH at vertex n_i to a currently Connected Component *CC*

Algorithm 1 aims at constructing the MST using the grid vertices representing the 3-D space candidate positions. Line 7 of Algorithm 1 search for the closest[5] two nodes in the initial set IS, which have the CHs and BS. If the closet two nodes are not adjacent on the 3-D grid (i.e. $P_c \leqslant \tau$), it adds at line 8 the minumum number of grid vertices on which the relays have to be placed to establish a path between these two nodes. After connecting the closest two nodes (i.e. establishinging a Connected Component *CC*), we iterativelly look for the next closest node that has to be connected to the *CC*. this has been achieved through lines 12–22 of Algorithm 1.

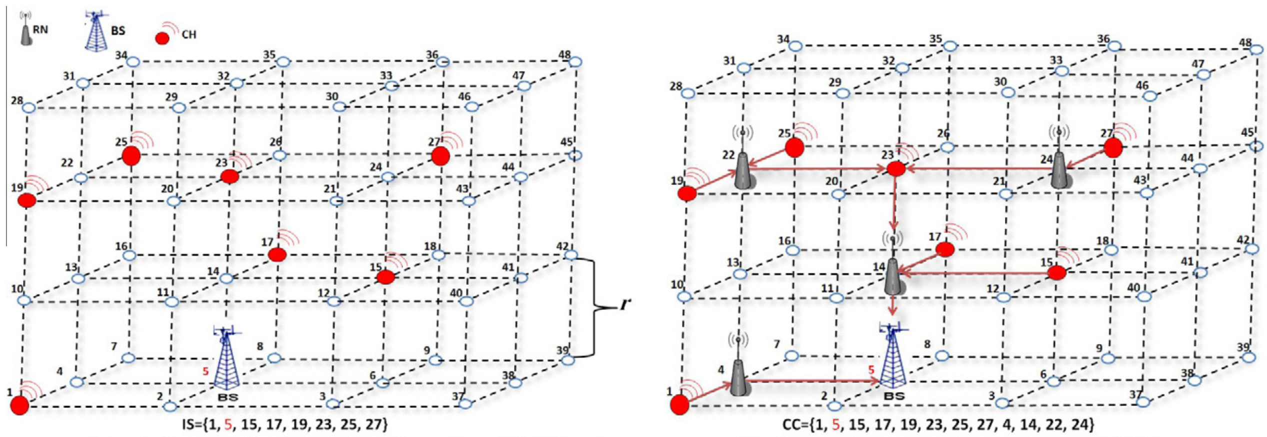
For more elaboration on placement of the FPRNs using Algorithm 1, consider the following example.

Example 1. Assume we have seven cluster heads preallocated with the base station on the grid vertices as in Fig. 7(a), then we seek the minimum number of relay nodes (=N_{MST}) required to connect these cluster heads with the base station as dipicted in Fig. 7(b). Positions of these N_{MST} relay nodes are determined by applying Algorithm 1. The algorithm first constructs the Initial Set *IS* consisting of cluster heads and base station coordinates. Then, a Connected Component *CC* set is initiated by the closest two nodes' coordinates in *IS*, which are cluster heads at vertices 15 and 17, in this example. These coordinates are then removed from *IS*. Obviously, by adding only one relay node at vertex 14, cluster heads at 15 and 17 become connected. Hence, coordinates of that relay node is added to *CC* and remaining number of nodes N_d in *IS* is set to 6. Now, we calculate M_i for the remaining nodes at vertices 1, 5, 19, 23, 25 and 27 in *IS*, which would have 2, 0, 2, 0, 1 and 1 candidate relay node coordinates, respectively. Since the set M_1, associated with the base station placed at vertex 5, has the smallest number of required coordinates (=0), we put M_1 in the set *SM* and *CC* becomes equal to {15, 17, 14, 5}. *M* and *IS* are then updated and N_d is decremented by 1. By repeating this process until N_d is equal to 0, we obtain the final connected component *CC* that is shown in Fig. 7(b) where relay nodes in this figure are the FPRNs of the wireless sensor network to be deployed. The deployed FPRNs with the cluster heads and the base station construct the network backbone. □

Connectivity of the Backbone *B* generated in this phase of the deployment is measured by considering *B* as a connected graph which has a Laplacian matrix $L(B)$ [14]. The Laplacian matrix is a two dimensional matrix that has -1 at the element (i,j), if there is a connection between nodes *i* and *j*. It has an integer positive number at the element (i,i) that represent number of edges connected to the node *i* (see Fig. 8). Given $L(B)$, the backbone connectivity (or algebraic connectivity) is mathematically measured by computing the second smallest eigenvalue λ_2. Where λ_2 indicates the minimum number of nodes and links whose removal would disconnect the graph *B* (see Fig. 8 for more elaboration on λ_2). By maximizing λ_2 of $L(B)$, we maximize the required number of nodes and communication links to disjoint (disconnect) paths in the network backbone. This is because of the proportional relationship between the value of λ_2 and the number of nodes/links which can cause network partitions according to Fig. 8. Hence, more reliable[6] environmental wireless sensor network can be achieved due to the ability to overcome significant topology changes caused by communication quality changes and node failures using tightly connected backbones. In order to *maximize* the backbone connectivity λ_2, extra relay nodes (SPRNs) are placed in the second phase of the O3DwLC strategy.

[5] In terms of the vertices count separating the two nodes.

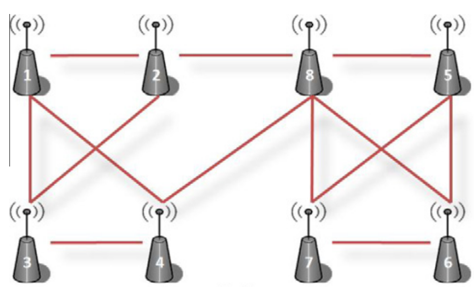
[6] Reliability here is defined by the existence of an operational path from all CHs to the BS even in the presence of nodes and links failure.



(a) Before applying Algorithm 1 (i.e. without FPRNs) and r is the grid edge length.

(b) After applying Algorithm 1 (i.e. with FPRNs).

Fig. 7. An example of FPRNs placement using Algorithm 1.



$$\begin{bmatrix} 3 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 3 & -1 & 0 & 0 & 0 & 0 & -1 \\ -1 & -1 & 3 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & -1 & 3 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 3 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 3 & -1 & -1 \\ 0 & 0 & 0 & 0 & -1 & -1 & 3 & -1 \\ 0 & -1 & 0 & -1 & -1 & -1 & -1 & 5 \end{bmatrix}$$

(a) **(b)**

Fig. 8. (a) A graph with 8 nodes and 13 links. The graph's connectivity characteristics are: one node to disconnect (removal of node 8), two links to disconnect (removal of links connecting node 8 to nodes 2 and 4), Laplacian matrix of this graph is shown in (b) and λ_2 of the matrix in (b) is equal to 0.6277. As λ_2 increases the node/link count required to partition the network increases.

Table 2
Notations used in the placement problem.

Notation	Description
α_i	A binary variable equals 1 when RN at vertex i in the 3-D grid is allocated and 0 otherwise
A_i	Incidence matrix that results by adding RN$_i$ in the 3-D grid; $A_i = [a_1, a_2, \ldots, a_m]$, where a_i is the vector that consists of n elements that can take a value of either 0, 1 or -1 and m is the total number of edges that is produced by adding RN$_i$. For example, if adding RN$_i$ will establish a connection between node 1 and 3, then 1st element is set to 1 and 3rd element is set to -1 and all of remaining elements are set to zeros
n	Summation of $N_{CH} + N_{MST} + 1$ (which is the total of CHs, FPRNs, and BS)
L_i	Initial Laplacian matrix produced by the allocated CHs, FPRNs, and the BS nodes
I_{nxn}	Identity matrix of size n by n

4.2. Second phase of the O3DwLC strategy

In this phase, we optimize positions of SPRNs such that λ_2 of the backbone generated in first phase is maximized with constraints on cost and lifetime. For simplicity, we start by maximizing λ_2 without lifetime constraints. Assume we have n_c grid vertices as candidate position for SPRNs. We want to choose the optimum N_{SPRN} relay nodes amongst these n_c relays with respect to connectivity; where N_{SPRN} is constrained by a cost budget. We can then formulate this optimization problem, with reference to Table 2, as

$$\max \lambda_2(L(\alpha)), \tag{6}$$

$$s.t. \quad \sum_{i=1}^{n_c} \alpha_i = N_{\mathrm{SPRN}}, \quad \alpha_i \in \{0, 1\},$$

where

$$L(\alpha) = L_i + \sum_{i=1}^{n_c} \alpha_i A_i A_i^T \tag{7}$$

However, an exhaustive search scheme is required to solve (6), which is computationally expensive, especially for the naturally large n_c values in large-scale environmental applications. This is due to the involved computations required for finding λ_2 for a large number $\left(= \left(\begin{matrix} nc \\ N_{\mathrm{SPRN}} \end{matrix} \right) \right)$ of Laplacian matrices. Therefore, we need a computationally efficient means to solve (6), in addition to more limited search space that reduces the value of n_c.

Taking advantage of the constructed network backbone in first phase, SPRNs may be placed on any grid vertex as long as it is within the probabilistic communication range of the largest number of CHs/FPRNs in B. This in turn can further finite the search space

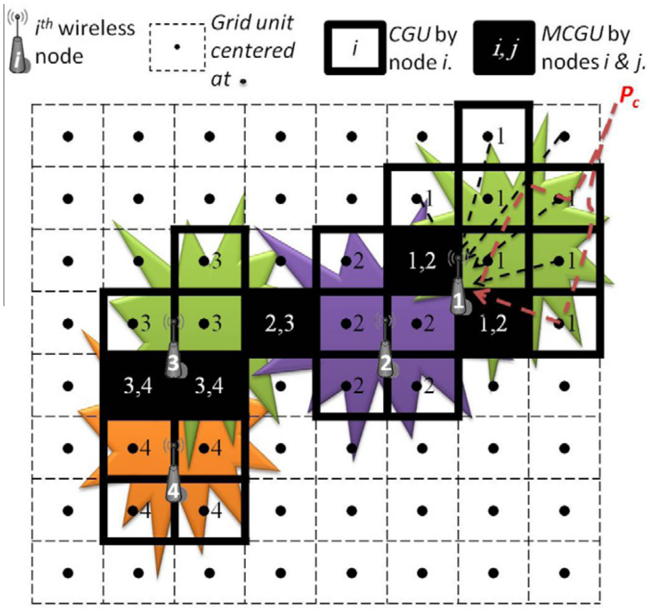


Fig. 9. An example of a maximum covered grid unit in 2-D plane. Numbers inside the bounded squares represent the node ID covering these squares.

without affecting the quality of the deployment plan. To explain our method of finding such a finite search space, we introduce the following definitions.

Ideal set definition: *A finite set of positions P is ideal iff it satisfies the following property:*

There exists an optimal[7]*placement of SPRNs 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 relays' positions are not including all of the grid vertices but a subset of these vertices that has the most potential to enhance the network connectivity. Moreover, since the computational complexity will be proportional to the cardinality of this ideal set, we should find a set with reasonably small size.

Covered grid unit (CGU) definition: *A covered grid unit α is a grid unit that has a connected center with at least one CH or FPRN. Let C (α) denote the subset of the CHs/FPRNs coordinates covering α.*

We assume that the considered virtual 3-D grid can have a building unit, called *grid unit*. For example, the *grid unit* of the 2-D grid shown in Fig. 9 is the small square drawn by dashed lines while in 3-D it will be a cube (in cubic grid models). Each *grid unit* is supposed to has a center of mass represented by its position coordinates (black dots in Fig. 9). We decide whether a *grid unit* is covered by a specific node (CH/FPRN), if the probabilistic connectivity P_c between the *grid unit* center and that node is greater than or equal to the aforementioned threshold τ.

Maximal covered grid unit (MCGU) definition: *A covered grid unit α is maximal if there is nocovered grid unit β, where $C(\alpha) \subseteq C(\beta)$.*

For more illustration, consider Fig. 9 in the 2-D plane. It shows four wireless nodes with respect to the grid units. Each wireless node has an arbitrary communication range. The covered grid units are bounded by solid lines, and the *maximal covered grid units (MCGUs)* are solid black squares. These MCGUs have the highest potential to place the SPRNs due to their ability to establish the highest number of new edges between already deployed FPRNs/CHs. Accordingly; we have to show that an ideal set can be derived from the set of MCGUs. Towards this end, we state the following Lemmas.

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

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

Lemma 2. *Finding a MCGU takes at most $(n-1)$ step, where n is number of nodes constructing the backbone B.*

Proof. By referring to the proof of Lemma 1, it is obvious that $|C(\alpha_x)| \leqslant n$, and $|C(\alpha)| < C(\alpha_1)| < |C(\alpha_2)| < \ldots < |C(\alpha_x) \leqslant n$; where $|C|$ represents the cardinality of the set C. Consequently, the process of finding the maximal covered grid unit α_x takes a finite number of steps less than or equal to $n-1$. □

Then, we introduce the following theorem.

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

Proof. To prove this theorem, it is sufficient to show that for any arbitrary placement Z we can construct an equivalent[8] placement Z' in which every SPRN is placed at a position in P. To do so, assume that in Z, a SPRN i is placed such that it is connected to a subset J of CHs/FPRNs. It is obvious that there exists a covered grid unit β, such that $J \subseteq C(\beta)$. From *Lemma 1*, there exist a MCGU α such that $C(\beta) \subseteq C(\alpha)$. In$Z'$, 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 CHs/FPRNs in J. By repeating for all SPRNs, we construct a placement Z' which is equivalent to Z, and thus Theorem 1 holds. □

In order to find all MCGUs, we need a data structure associated with each *grid unit* to store coordinates and total number of CHs/FPRNs covering the grid unit. We represent this data structure by the *covered grid unit* set $C(i)$, where i is the center of the grid unit. By computing $C(i)$, $\forall i \in V$,we can test whether a covered grid unit centered at i is maximal or not by searching for a set that has at least all elements of $C(i)$. In the following, Algorithm 2 establishes the data structures in $O(n)$. It associates each grid unit center with its total covering nodes. In line 9 of Algorithm 2, we computes the probability of the grid unit center i being connected with each backbone node individually based on Eq. (2). This is repeated by lines 5–15 until all probabilities between the grid units' centers and all backbone nodes are computed and the data structure is formed. Algorithm 3 tests whether a *covered grid unit* is maximal or not. In line 11 of Algorithm 3, we search for any set (other than C_i) in C that has the same backbon nodes which cover the grid center i. if such a set is found, Algorithm 3 returns false, otherwise it returns true; meaning that the set C_i is maximal. Algorithm 4 uses Algorithms 2 and 3 to construct an ideal set P by finding all MCGUs. The overall complexity of Algorithm 4 is $O(n\log n)$.

[7] Optimal in terms of connectivity.

[8] Equivalent in terms of the covering CHs/FPRNs. In other words, the placement of a SPRN at position i, within the communication range of the CHs x and y, is equivalent to the placement of the same SPRN node at position j within the communication range of the CHs x, y and z.

Algorithm 2. Build up grid units data structure.

```
Function FindGridUnitCoverage (B: Backbone constructed
  by CHs & FPRNs)
Input:
A set B of the CHs and FPRNs nodes' coordinates.
Begin
foreach grid unit center i do
C(i) := ∅;
foreach CH/FPRN j do
Compute Pc(i,j);
If Pc(i,j) ⩾ τ
C(i) := j ∪ C(i);
endif
endfor
endfor
End
```

Algorithm 3. Testing whether a grid unit set $C(i)$ is maximal or not.

```
Function Maximal (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 MCGU and False otherwise.
Begin
If C(i) := ∅ do
return False;
endif
Search for a set C′ such that C(i) ⊆ C′.
If C′ := ∅ do
return True;
else
return False;
endif
End
```

Algorithm 4. Finding all *Maximal Covered Grid Units* MCGUs.

```
Function FindMCGUs (B: Backbone constructed by CHs &
  FPRNs)
Input:
A set B of the CHs and FPRNs nodes' coordinates.
Output:
A set P that contains one position from every MCGU.
Begin
P := ∅;
FindGridUnitCoverage(B);
foreachC(i) do
If Maximal(C(i), all non-empty grid unit sets) do
P := {i} ∪ P;
endif
endfor
End
```

Once we obtain the set P which contains one position (grid vertex coordinates) from each MCGU, the search space of the problem formulated in (6) becomes much more limited.

In order to efficiently solve the optimization problem in (6), we reformulate it as a standard semi-definite program (SDP) optimization problem [7,14], which can be solved using any standard SDP solver. By relaxing the Boolean constraint $\alpha \in \{0,1\}$ to be linear constraint$\alpha \in [0,1]$, we can represent the problem in (6) as

$$\max \lambda_2(L(\alpha)), \qquad (8)$$
$$s.t. \quad \sum_{i=1}^{n_c} \alpha_i = N_{SPRN}, \quad 0 \leqslant \alpha_i \leqslant 1,$$

The optimization problem in (8) is convex with linear constraint [7]. Thereby we introduce the following lemma.

Theorem 2. *The optimization problem in (8) is mathematically equivalent to the following SDP optimization problem*

$$\max S \qquad (9)$$
$$s.t. \quad S\left(\boldsymbol{I}_{nxn} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right) \preccurlyeq L(\alpha), \quad \sum_{i=1}^{n_c} \alpha_i = N_{\mathrm{SPRN}}, \quad 0 \leqslant \alpha_i \leqslant 1,$$

where S is a scalar variable and denotes the positive semi-definiteness (i.e. all eigenvalues of the matrix are greater than or equal to zero).

Proof. Let $V \in \mathbf{R^n}$ be the corresponding eigenvector of $\lambda_2(L(\alpha))$. Thus, $\mathbf{1}^T V = 0$, and $\|V\| = 1$. Since,

$$L(\alpha)V = \lambda_2 V. \qquad (10)$$

Hence,

$$V^T L(\alpha)V = \lambda_2 V^T V = \lambda_2. \qquad (11)$$
$$\Rightarrow \lambda_2(L(\alpha)) = inf_V\{V^T L(\alpha)V | \mathbf{1}^T V = 0, \text{ and } \|V\| = 1\}(12)\lambda_2(L(\alpha)). \qquad (12)$$

Let,

$$L'(\alpha) = L(\alpha) - S\left(\boldsymbol{I_{nxn}} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right). \qquad (13)$$

Thus for any $\|V\|_{\mathrm{nx1}}$ where $\mathbf{1}^T V = 0$, *and* $\|V\| = 1$, we get

$$V^T L'(\alpha)V = V^T L(\alpha)V - S\left(\boldsymbol{V^T I_{nxn} V} - \frac{1}{n}(\boldsymbol{V^T 1})(\boldsymbol{1^T V})\right)$$
$$= V^T L(\alpha)V - S. \qquad (14)$$

Hence, for $L'(\alpha)$ to be positive semi-definite, the maximum value of S should be

$$S = inf_V\{V^T L(\alpha)V | \mathbf{1}^T V = 0, \text{ and } \|V\| = 1\}. \qquad (15)$$

From (12) and (15),

$$S = \lambda_2(L(\alpha)). \qquad (16)$$

Therefore, maximizing S in (9) is equivalent to maximizing $\lambda_2(L(\alpha))$ in (8) if the constraints are satisfied. □

In order to add environmental lifetime constraints to Eq. (9), let the backbone (generated in first phase) be operational for initial number of rounds equal to *IRs*. Assume adding one relay node of the SPRNs would prolong the network lifetime by extra rounds ER_i. Then, to guarantee that the network will stay operational for a minimum number of required rounds *RLT*, the total extra and initial rounds must be greater than or equal to *RLT* as elaborated in the following:

$$-\sum_{i=1}^{n_c} ER_i \leqslant (IRs - RLT). \qquad (17)$$

Since we are using the cutoff criterion of the *environmental lifetime* definition in calculating both ER_i and *IRs*, inequality (17) represents a more environment-specific lifetime constraint in the O3DwLC strategy.

From (9) and (17), SPRNs positions that maximize λ_2 with constraints on lifetime and cost are found by solving the following[9]:

$$\max S, \tag{18}$$
$$s.t.\quad S\left(\boldsymbol{I}_{nxn} - \frac{1}{n}\boldsymbol{1}\boldsymbol{1}^T\right) \leqslant L(\alpha), \quad \sum_{i=1}^{n_c} \alpha_i \leqslant N_{\mathrm{SPRN}},$$
$$-\sum_{i=1}^{n_c} ER_i\alpha_i \leqslant (IRs - RLT), \quad 0 \leqslant \alpha_i \leqslant 1,$$

In the following, Algorithm 5 summarizes the second phase deployment proposed in this section where the search space is limited to n_c positions for grid vertices within the ideal set P.

Algorithm 5. SPRNs deployment

1. **Function SPRNs** (B: Backbone constructed by CHs, FPRNs & BS, *P*)
2. **Input:**
3. A set B of the CHs, FPRNs and BS nodes' coordinates.
4. An ideal set *P* of n_c candidate positions for the SPRNs.
5. **Output:**
6. A set SP of the SPRNs coordinates maximizing connectivity of B with practical lifetime and cost constraints
7. **begin**
8. $\mathbf{L_i}$ = Laplacian matrix of B
9. **IRs** = number of rounds B can stay operational for
10. **for**(i = 1; i < n_c; i++)
11. $\mathbf{A_i}$ = coefficient matrix corresponding to vertex i on the grid
12. $\mathbf{ER_i}$ = extra rounds achieved by allocating RN at vertex i
13. **end**
14. **SP** = Solution of SDP in (18)
15. **End**

For more ellaboration on Algorithm 5, consider the following example.

Example 2. Assume we have up to two extra relay nodes (SPRNs) to maximize connectivity of the backbone generated in Fig. 7(b) and ensure at least 20 rounds the network can stay operational for. In this case, $N_{SPRN} = 2, n = 12, and\, RLT = 20$. We start by computing the ideal set *P* to specify our search space in this problem using Algorithm 4. Afterward, we calculate the initial Laplacian matrix L_i associated with the backbone to be used in Eq. (7), in addition to computing initial rounds ($IRs = 10$) the backbone can stay operational for. With reference to Table 2, we set α_i to 1 and calculate A_i and extra rounds ER_i for each element *i* in the ideal set *P*. Notice that *P* in this example is the set of vertices 10, 13, 16, 2, 8, 20, 26, 6, and 18 in Fig. 7(b), assuming that only nodes placed on adjacent vertices are connected. Now we solve the SDP in (18) for this specific example. As a result, the highest two values of λ_2 (i.e. network connectivity with constraints on cost and lifetime) are associated with vertices 10 and 26. By allocating the two SPRNs at these two vertices we guarantee the network lifetime to be at least 20 rounds, in addition to maximizing the backbone connectivity, as well. For instance, we can see how removal of a single node such as FPRNs at vertex 4 or 14, in Fig. 7(b), can cause a network partition. While using the SPRNs, allocated at vertices 10 and 26, at least two nodes removal is required to cause the network partition. □

Finally, based on the output of Algorithms 1 and 5, locations of relay nodes (FPRNs and SPRNs) are determined optimally in terms of maximum connectivity and limited cost budget, in addition to practical lifetime considerations. This can be easily proven given that the two solutions achieved by Algorithms 1 and 5 are optimal and independent. We remark that this two-phase solution can be easily extended to consider other constraints such as coverage, data fidelity, fault-tolerance, etc. It is also important to notice that Algorithms 1 and 5 are computationally efficient in practice with complexity of $O(n)$.

5. Performance evaluation

In this section, we evaluate the performance of our proposed O3DwLC strategy under harsh environmental circumstances, where numerous probabilities of node failure and isolation are considered and 3-D setup is required. We compare our strategy to an efficient deployment strategy, called the Shortest Path 3-D grid deployment (SP3D). The SP3D strategy is usually used in environmental applications such as forest fire detection and soil experiments [29,31]. Moreover, SP3D strategy is used as a baseline in this research due to its efficiency in maintaining a predefined lifetime and choosing the minimum number of relay nodes required in constructing the network backbone. In SP3D, Algorithm 1 is used to construct the network backbone by allocating the minimum number of relay nodes on 3-D grid vertices. These relay nodes connect the pre-allocated cluster heads with the base station. Then, extra relay nodes are densely distributed near to the network backbone devices in order to enhance their connectivity. Both O3DwLC and SP3D strategies are evaluated and compared using three different metrics:

1. *Backbone Connectivity* (λ_2): this criterion reflects deployed network reliability under harsh environmental characteristics and ability to prolong lifetime. It indicates efficiency of the designed wireless sensor network.
2. *Number of CHs/RNs:* this indicates the system cost effectiveness in harsh environments.
3. *Number of rounds:* this is a measurement of the total rounds the deployed network can stay operational for. It reflects efficiency of the estimated wireless sensor network lifetime.

Two main parameters are used in this comparison: Probability of Node Failure (PNF) and Probability of Disconnected (isolated) Nodes (PDN). PNF is the probability of physical damage for the deployed node. PDN is the probability of a node to be disconnected while it still has enough energy to communicate with the base station. We chose these parameters as they are key factors in reflecting harshness of the monitored site in terms of weak signal reception and physical node damage.

5.1. Simulation model

The O3DwLC and SP3D strategies are executed on 500 randomly generated WSNs hierarchical graph topologies in order to get statistically stable results. For each topology, we apply a random node/link failure and performance metrics are computed accordingly. Dimensions of the 3-D deployment space are $700 \times 700 \times 200$ (m^3). Twenty irredundant cluster heads (i.e. responsible for different sensor nodes) in addition to one base station are randomly placed on 3-D cubic grid vertices using a Linear Congruential random number generator. We assume a predefined fixed time schedule for traffic generation at the cluster heads. Positions of relay nodes are found by applying the O3DwLC and the SP3D deployment strategies. We assume that each wireless sensor network is required to be operational for at least 20 rounds (lifetime constraint) using at most 60 relay nodes (cost constraint).

[9] SDPA-M Matlab package can be used to solve (18).

Table 3
Parameters of the simulated WSN.

Parameter	Value	Parameter	Value
RLT	20 (round)	N_{SPRN}	0–60 (relay node)
Total grid units	98	PNF	0–60%
PDN	0–60%	N_{CH}	20

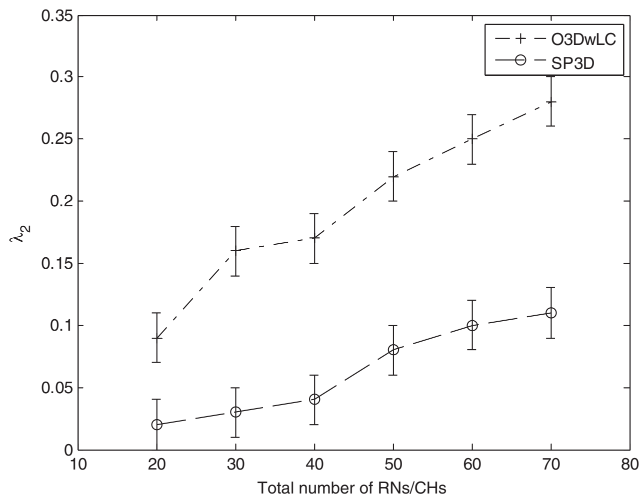


Fig. 10. Connectivity vs. the deployed nodes' count.

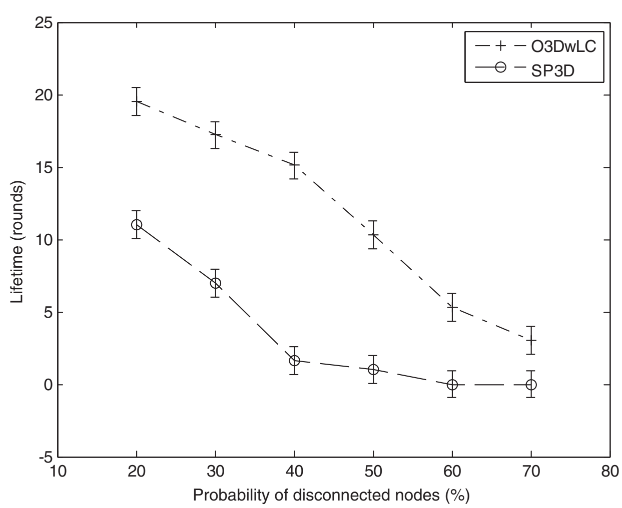


Fig. 11. Lifetime vs. PDN.

Based on experimental measurements taken in a site of dense trees [28], we set our system model variables to be as described in Tables 1 and 3. However, the n_c variable, which represent the search space size for the optimization problem in (18), will have different values varying from one deployment to another based on the locations of the backbone nodes reached by Algorithm 1 and their probabilistic communication ranges. Thus, the formulated sets of CGUs and MCGUs will vary from one deployment instance to another after applying Algorithms 2–4 on the resulted network backbone in the first phase of the O3DwLC strategy. Similarly, the values of N_{MST} and *IRS* are assigned based on the results of the first phase deployment, and consequently, they vary from one instance to another. We assume fixed and equal transmission ranges to simplify the presentation of results, in addition to applying identical grid edge lengths (=100 m). Nevertheless, the same

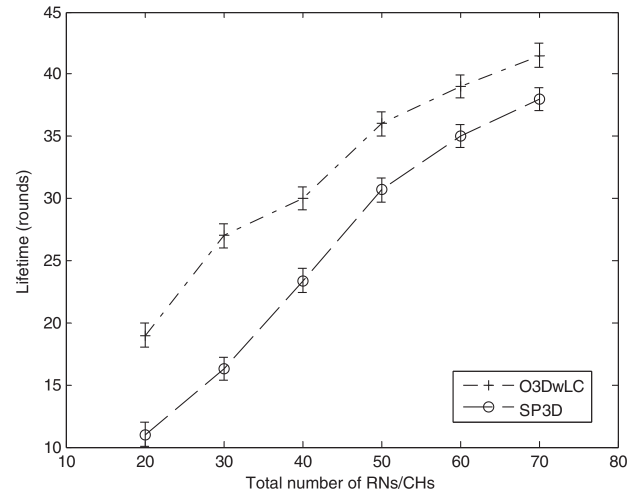


Fig. 12. Lifetime vs. number of nodes under PDN = 0.2 and RLT = 20.

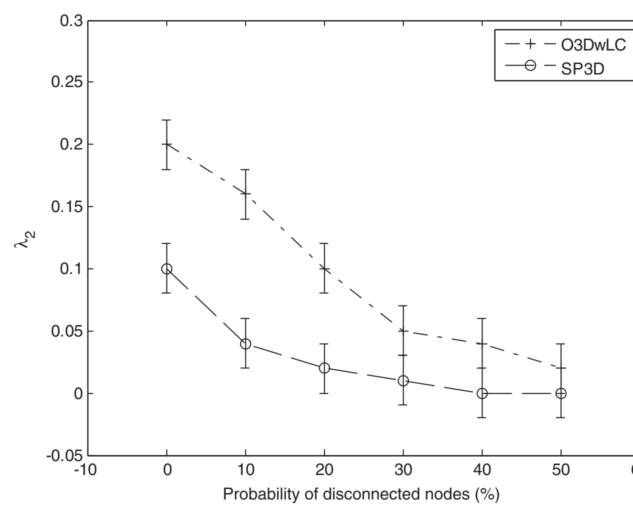


Fig. 13. Connectivity vs. probability of disconnected CHs/RNs.

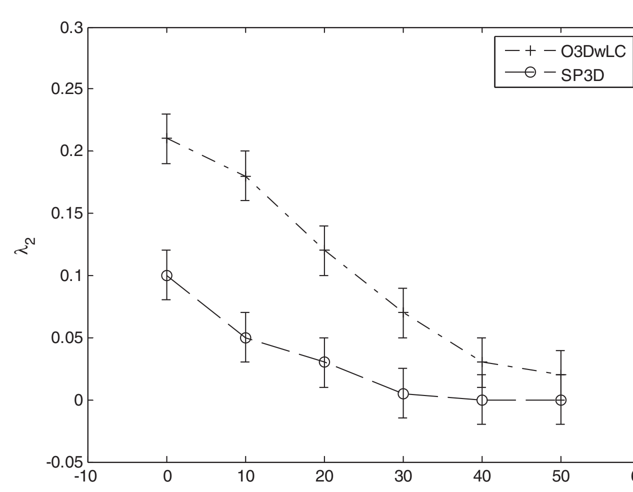


Fig. 14. WSN connectivity vs. probability of CHs/RNs failure.

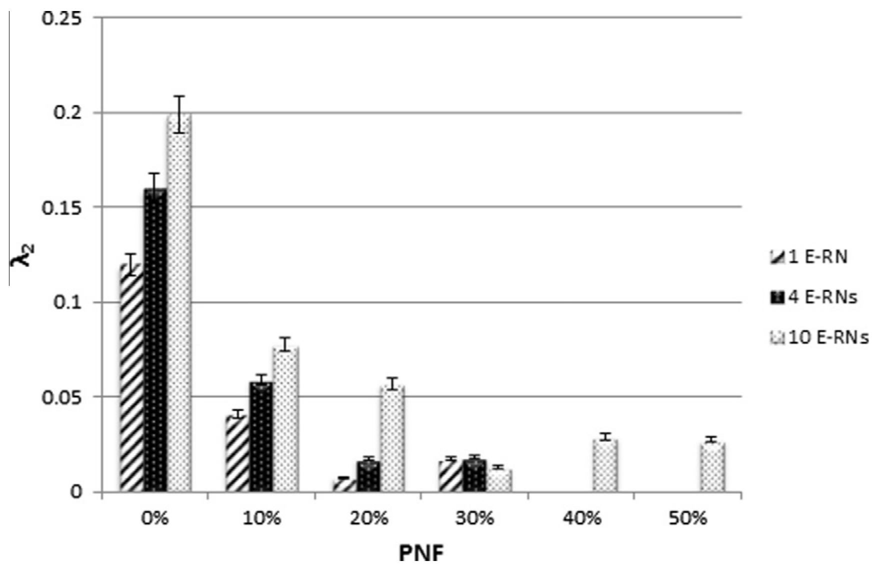


Fig. 15. Indicating SPRNs count required to overcome specific PNF value.

simulation parameters are applicable in case of different transmission ranges varying from one device to another with unequal grid edge lengths. For comparison purposes, we simulate the lifetime and connectivity of different relay nodes counts and varying PNF and PDN (0–60%) for both strategies, O3DwLC and SP3D. As we aforementioned, each simulation experiment is repeated 500 times and the average results hold a confidence interval no more than 5% of the average (over 500 runs) at a 95% confidence level.

5.2. Simulation results

While the O3DwLC strategy optimizes locations of second phase relay nodes in order to achieve the highest connectivity (λ_2), SP3D attempts to find locations of extra relay nodes (deployed after constructing the optimized backbone) by distributing them on 3-D grid vertices that are more likely to increase the network connectivity as we described above. It is expected from O3DwLC to outperform the SP3D in terms of connectivity and total number of nodes as shown in Fig. 10. Fig. 10 presents the average λ_2 for both strategies using different total numbers of RNs nodes, where number of cluster heads is fixed to 20 to see the effect of relay node placement, and PDN = 0.2. It is clear how an increment in the deployed RNs leads to increment in connectivity even in the presence of 20% disconnected nodes using O3DwLC strategy. Moreover, Fig. 10 shows how efficient are the networks generated by O3DwLC in terms of the utilized RNs count (and thus, the cost) to achieve a specific connectivity requirement. For instance, using 30 nodes only, O3DwLC strategy achieves a connectivity value higher than the connectivity value achieved by SP3D using 70 nodes which indicates a higher save in terms of the network cost.

We also investigate the effect of PDN, which varies from 0–60%, on lifetime using both strategies as shown in Fig. 11. We select the total number of nodes in this comparison to be 40 (20 CHs and 20 RNs) to avoid extreme cases where total number of RNs is too high or too low. From Fig. 11, we can see how the WSNs generated by O3DwLC can stay operational for longer time than the SP3D which indicates high reliability, where there exist at least one operational path from each cluster head to the base station even in the presence of 60% PDN [1]. From Figs. 10 and 12 we also observe that increasing the connectivity value of the WSN will increase its lifetime since both figures are simulating the same WSNs with the same total number of nodes.

In Fig. 12, we examine the effect of the lifetime constraint on the O3DwLC and SP3D strategies when PDN = 20%. Under that PDN, O3DwLC is still much better in terms of the total rounds a network can stay operational for. This is a very attractive feature in outdoor monitoring. Not surprisingly, the difference in lifetime of the WSNs generated using both strategies decreases as the total number of relay nodes increases due to node density increment which makes the deployed networks tightly connected and harder to partition.

Figs. 13 and 14 show how O3DwLC strategy outperforms SP3D strategy under different PDN and PNF values, respectively. This supports our O3DwLC deployment strategy efficiency in terms of connectivity. Wireless sensor networks generated by O3DwLC strategy stays connected even under PDN = PNF = 50%. This is another desired and required performance issue in harsh outdoor environmental applications. However, as the PNF/PDN values achieve a specific level where the available number of functional nodes cannot tolerate the failure, the WSN connectivity decreases dramatically and network partitions occurs. This explains the prominent degradation in WSN connectivity even when O3DwLC strategy is utilized.

We remark that choosing an appropriate value of N_{SPRN} is highly dependent on the probability of node failure. Fig. 15 shows the effect of PNF on the choice of N_{SPRN}. For low values of PNF only few Extra Relay Nodes (E-RNs) are needed. On the other hand, at least 10 E-RNs are needed to guarantee connectivity in environments with a 50% PNF. Considering such percentage of failure in the monitored site during the early stages of the deployment plan would have a great effect on the network performance in practice.

6. Conclusion

In this paper, we explored the problem of relays deployment in WSNs applied in 3D outdoor environmental applications; aiming at maximizing network connectivity with constraints on lifetime and cost. Such deployment problem has been shown to be NP-hard. Finding near optimal solutions is also an NP-hard. To address this complexity, we propose an efficient two-phase relay node deployment in 3D space using minimum spanning tree and semi-definite programming. The first phase is using the minimum spanning tree to setup a connected network backbone with the minimum number of relay nodes for cost efficiency. In the second phase, we found a set of a relatively small count of candidate positions. And we used the semi-definite programming to optimize the relay nodes placement on these positions to achieve the maximum backbone connectivity for guaranteed lifetime period within a limited cost budget. Towards more practical solution, application-specific signal propagation and lifetime models were considered, in addition to limiting the huge search space of the targeted deployment problem.

The signal propagation model provided more realistic communication properties between the deployed nodes in order to precisely describe their ability to communicate between each other. As for the lifetime model, several lifetime definitions in the literature were discussed and compared based on practical metrics and parameters to reflect the appropriateness of these definitions in environmental applications. The extensive simulation results, obtained under harsh operational conditions, indicated that the proposed two-phase strategy can provide tightly-connected networks and practically-guaranteed lifetime for environmental applications. Moreover, deployment strategy and results presented in this paper can provide a tangible guide for network provisioning in large-scale environmental applications which require 3-D setups. In addition, they are applicable for different grid shapes and environment characteristics (e.g. various signal attenuation and path loss levels).

Future work would investigate optimal deployment problem in further environment monitoring scenarios, where a subset of the relay nodes may have the mobility feature to repair connectivity and prolong network lifetime. Also, of practical interest is the node placement under varying transmission range and/or different power supply from one node to another for more energy-efficient solutions.

Acknowledgment

This work has been supported by the Ontario Centers of Excellence (OCE), Ontario, Canada.

References

[1] A. Abbasi, M. Younis, K. Akkaya, Movement-assisted connectivity restoration in wireless sensor and actor networks, IEEE Transactions on Parallel Distributed Systems 20 (9) (2009) 1366–1379.
[3] I. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, A survey on sensor networks, IEEE Communications Magazine 40 (8) (2002) 102–114.
[5] A. Bari, A. Jaekel, S. Bandyopadhyay, Optimal placement of relay nodes in two-tiered, fault tolerant sensor networks, in: Proceedings of the IEEE International Conference on Computers and Communications (ISCC), Aveiro, 2007, pp. 159–164.
[6] M. Bhardwaj, A. Chandrakasan, T. Garnett, Upper bounds on the lifetime of sensor networks, in: Proceedings of the IEEE International Conference on Communications (ICC), St. Petersburg, 2001, pp. 785–790.
[7] S. Boyd, Convex optimization of graph laplacian eigenvalues, in: Proceedings of the International Congress of Mathematicians, vol. 3, no. 63, November 2006, pp. 1311–1319.
[8] A. Cerpa, D. Estrin, Ascent: adaptive self-configuring sensor networks topologies, IEEE Transactions on Mobile Computing 3 (3) (2004) 272–285.
[9] Y. Chen, Q. Zhao, On the lifetime of wireless sensor networks, IEEE Communications Letters 9 (11) (2005) 976–978.
[11] X. Cheng, D. Du, L. Wang, B. Xu, Relay sensor placement in wireless sensor networks, Journal of Wireless Networks 14 (3) (2008) 347–355.
[12] C. Decayeux, D. Seme, A new model for 3-D cellular mobile networks, in: Proceedings of the International Symposium Parallel and Distributed Processing (ISPDC), Cork, Irlande, 2004.
[13] A. Efrat, S. Fekete, P. Gaddehosur, J. Mitchell, V. Polishchuk, J. Suomela, Improved approximation algorithms for relay placement, in: Proceedings of the 16th Annual European Symposium on Algorithms, Karlsruhe, Germany, 2008, pp. 356–367.
[14] A. Ghosh, S. Boyd, Growing well-connected graphs, in: Proceedings of the IEEE Conference on Decision and Control, San Diego, CA, 2006, pp. 6605–6611.
[15] L. Hoesel, T. Nieberg, J. Wu, P. Havinga, Prolonging the lifetime of wireless sensor networks by cross layer interaction, IEEE Transactions on Wireless Communications 11 (6) (2004) 78–86.
[17] M. Ibnkahla (Ed.), Adaptation and Cross Layer Design in Wireless Networks, CRC, Boca Raton, 2008.
[18] M. Ishizuka, M. Aida, Performance study of node placement in sensor networks, in: Proceedings of the International Conference on Distributed Computing Systems Workshops (Icdcsw), Tokyo, 2004, pp. 598–603.
[19] A. Kashyap, S. Khuller, M. Shayman, Relay placement for higher order connectivity in wireless sensor networks, in: Proceedings of the IEEE Conference on Computer Communications (INFOCOM), Barcelona, Spain, 2006, pp. 1–12.
[20] S. Lee, M. Younis, Optimized relay placement to federate segments in wireless sensor networks, IEEE Transactions on Selected Areas in Communications 28 (5) (2010) 742–752.
[21] J. Lee, T. Kwon, J. Song, Group connectivity model for industrial wireless sensor networks, IEEE Transactions on Industrial Electronics 57 (5) (2010) 1835–1844.
[22] N. Li, J.C. Hou, Improving connectivity of wireless ad hoc networks, in: Proceedings of the IEEE International Conference on Mobile and Ubiquitous Systems: Networking and Services (MobiQuitous), San Diego, CA, 2005, pp. 314–324.
[23] H. Liu, P. Wan, X. Jia, Fault-tolerant relay node placement in wireless sensor networks, Lecture Notes in Computer Science (LNCS) 3595 (2005) 230–239.
[24] E. Lloyd, G. Xue, Relay node placement in wireless sensor networks, IEEE Transactions on Computers 56 (1) (2007) 134–138.
[25] S. Misra, S. Hong, G. Xue, J. Tang, Constrained relay node placement in wireless sensor networks to meet connectivity and survivability requirements, in: Proceedings of the IEEE Conference on Computer Communications (INFOCOM), Phoenix, AZ, 2008, pp. 281–285.
[26] C. Ortiz, J. Puig, C. Palau, M. Esteve, 3D wireless sensor network modeling and simulation, in: Proceedings of the IEEE Conference on Sensor Technologies and Applications (SensorComm), Valencia, Spain, 2007, pp. 307–312.
[28] J. Rodrigues, S. Fraiha, H. Gomes, G. Cavalcante, A. de Freitas, G. de Carvalho, Channel propagation model for mobile network project in densely arboreous environments, Journal of Microwaves and Optoelectronics 6 (1) (2007) 189–206.
[29] A. Singh, M.A. Batalin, V. Chen, M. Stealey, B. Jordan, J.C. Fisher, T.C. Harmon, M.H. Hansen, W.J. Kaiser, Autonomous robotic sensing experiments at san Joaquin river, in: Proceedings of the IEEE International Conference of Robotics and Automation (ICRA), Roma, 2007, pp. 4987–4993.
[30] W. Song, R. Huang, M. Xu, B. Shirazi, R. LaHusen, Design and deployment of sensor network for real-time high-fidelity volcano monitoring, IEEE Transactions On Parallel and Distributed Systems 21 (11) (2010) 1658–1674.
[31] B. Son, Y. Her, J. Kim, A design and implementation of forest-fires surveillance system based on wireless sensor networks for South Korea mountains, International Journal of Computer Science and Network Security 6 (9) (2006) 124–130.
[32] H. Tan, Maximizing network lifetime in energy-constrained wireless sensor network, in: Proceedings of the ACM International Wireless Communications and Mobile Computing Conference (IWCMC), Vancouver, BC, 2006, pp. 1091–1096.
[33] J. Tang, B. Hao, A. Sen, Relay node placement in large scale wireless sensor networks, Computer Communication. 29 (4) (2006) 490–501.
[34] G. Tolle, J. Polastre, R. Szewczyk, D. Culler, A macroscope in the redwoods, in: Proceedings of the ACM Conference on Embedded Networked Sensor Systems (SenSys), San Diego, 2005, pp. 51–63.
[35] K. Xu, H. Hassanein, G. Takahara, Q. Wang, Relay node deployment strategies in heterogeneous wireless sensor networks, IEEE Transactions on Mobile Computing 9 (2) (2010) 145–159.
[37] M. Younis, K. Akkaya, Strategies and techniques for node placement in wireless sensor networks: a survey, Elsevier Ad Hoc Network Journal 6 (4) (2008) 621–655.

Further Reading

[2] H.M.F. AboElFotoh, S.S. Iyengar, K. Chakrabarty, Computing reliability and message delay for cooperative wireless distributed sensor networks subject to random failures, IEEE Transactions on Reliability 54 (1) (2005) 145–155.
[4] S.N. Alam, Z. Haas, Coverage and connectivity in three-dimensional networks, in: Proceedings of the ACM International Conference on Mobile Computing and Networking (MobiCom), Los Angeles, CA, 2006, pp. 346–357.
[10] P. Cheng, C. Chuah, X. Liu, Energy-aware node placement in wireless sensor networks, in: Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM), Dallas, Texas, 2004, pp. 3210–3214.
[16] Y. Hou, Y. Shi, H. Sherali, S.F. Midkiff, On energy provisioning and relay node placement for wireless sensor network, IEEE Transactions on Wireless Communications 4 (5) (2005) 2579–2590.
[27] V. Ravelomanana, Extremal properties of three-dimensional sensor networks with applications, IEEE Transactions on Mobile Computing 3 (3) (2004) 246–257.
[36] W. Ye, J. Heidemann, D. Estrin, Medium access control with coordinated adaptive sleeping for wireless sensor networks, IEEE/ACM Transactions on Networks 12 (3) (2004) 493–506.