Optimal Placement of Relay Nodes Over Limited Positions in Wireless Sensor Networks

Miloud Bagaa, Ali Chelli, Djamel Djenouri, Tarik Taleb, Ilangko Balasingham and Kimmo Kansanen

Abstract

This paper tackles the challenge of optimally placing relay nodes (RNs) in wireless sensor networks (WSN) given a limited set of positions. The proposed solution consists in i) the usage of a realistic physical layer model based on a Rayleigh block-fading channel, ii) the calculation of the signal-to-interference-plus-noise ratio (SINR) considering the path loss, fast fading, and interference, and iii) the usage of a weighted communication graph drawn based on outage probabilities determined from the calculated SINR for every communication link. Overall, the proposed solution aims for minimizing the outage probabilities when constructing the routing tree, by adding a minimum number of RNs that guarantee connectivity. In comparison to the state-of-the art solutions, the conducted simulations reveal that the proposed solution exhibits highly encouraging results at a reasonable cost in terms of the number of added RNs. The gain is proved high in terms of extending the network lifetime, reducing the end-to-end- delay, and increasing the goodput.

Index Terms

M. Bagaa and T. Taleb are with Communications and Networking Department, Aalto University, Finland (emails: miloud.bagaa@aalto.fi and talebtarik@ieee.org).

A. Chelli, I. Balasingham and K. Kansanen are with the Department of Electronics and Telecommunications, Norwegian University of Science and Technology (NTNU), Trondheim, Norway (emails: {ali.chelli, ilangkob, kimmo.kansanen}@iet.ntnu.no).

D. Djenouri is with the DTISI, CERIST Research Center, Algiers, ALGERIA (e-mail: ddjenouri@acm.org).

Relay node placement, wireless sensor network, connectivity, SINR model.

I. INTRODUCTION

A wireless sensor network (WSN) is a collection of sensor nodes (SNs) that are deployed to sense the environment and transmit sensed information, possibly over multiple hops, to the base stations (BSs). As SNs are generally limited in resources and power supply, additional relay nodes (RNs) are deployed to forward the collected data from the SNs to the BSs and thus extending the network lifetime of the SNs. Those dedicated RNs are either power rich nodes [1], or have capacity to harvest energy from the environment [2].

While a lot of research has been conducted on the topic of relay node placement [1], relatively little has been done for *constrained* relay placement. We are interested in the latter case, where the potential positions for RNs placement within the deployment area are limited. This is more realistic for most application scenarios than unconstrained models that suppose possible placement at any position. From the routing structure perspective, relay node placement models can be classified into two categories, single-tiered vs. two-tiered [1]. In *single-tiered* relay node placement, both SNs and RNs forward packets, while in *two-tiered* model only RNs forward packets, and the SNs use is limited to data collection and first-mile transmission to a RN (or directly to a BS if possible). We are interested in the second model, which is more power efficient and permits to preserve the limited resources of SNs. As proven in [1], determining the optimal solution for the two-tiered constrained relay node placement is NP-hard, and thus heuristics for solving this problem have to be developed. We propose in this paper a constrained RNs deployment scheme while considering a realistic physical model which takes into account propagation channel conditions.

The physical model is based on a Rayleigh block-fading channel and captures path loss,

fast fading, and interference to derive the signal-to-interference-plus-noise ratio (SINR). The cumulative density function (CDF) of the latter is calculated and used to determine the outage probability–for every communication link and in each directions–as function of the receiver sensitivity threshold. The outage probability is then used to calculate the weights of the communication graph edges. This graph will be used to determine the selected (activated) positions where to put relays. Therefore, instead of focusing just on minimizing the number of added relays that assure a certain connectivity (which has been largely treated in the literature, e.g, [1], [3]), the aim here is to minimize the outage probabilities by deploying a minimum number of relays in a set of candidate positions. Minimizing this outage probability allows to extend the network lifetime and enhance the quality of service (QoS). To our knowledge, this paper is the first that considers physical layer parameters in the constrained relay node placement problem. The proposed solution has been analyzed, and compared to some state-of-the-art solutions. The results show clear improvement in all performance metrics, including, goodput, end-to-end delay, and network life time, at a reasonable increase in the cost (number of added RNs).

The rest of the paper is organized as follows. An overview on the literature of the relay node placement problem is provided in Section II. The problem formulation and the network model are presented in Section III, followed by the description of the relay node placement algorithm in Section IV. Section V provides theoretical analysis of the physical model, while the simulation results are presented in Section VI. Finally, Section VII concludes the paper.

II. RELATED WORK

This section focuses on the review of the prior work on relay node placement, which can be classified in two main categories, namely, single-tiered and two-tiered relay node placement. For the first category, a sensor node can be used by other nodes in the network for data forwarding,

whereas for the second one, a sensor node only sends its own data but cannot be utilised by other nodes in the network to forward their data.

In the literature, many research works have studied WSN with single-tiered topology [4]–[7]. The authors of [4] considers the relay node placement problem such that a survivability requirement is achieved. In this problem, the aim is to determine the location of the minimum number of relay nodes such that each sensor is connected to a base station through several node-disjoint paths, which provides fault tolerance in case of node failure. The authors of [4] propose an algorithm that deploys a minimum number of nodes such that a predefined level of survivability is achieved. The solution reached by the algorithm, which has a single-tiered topology, is proven to be within a constant factor from the global optimum. The connectivity problem in WSN is addressed in [5]. In the connectivity problem, the target is to find the minimum number of relays such that each sensor is connected with a BS. The authors of [5] aim to find a tradeoff between performance in terms of network lifetime and cost. The problem is modelled by a Steiner tree with minimum number of Steiner points and bounded edge-length. Two approximation algorithms are proposed and their performance is analysed. [6] is one of the first work that studied the Steiner tree problem with minimum number of Steiner points and bounded edge-length and proved that this problem is NP-complete. The authors present a polynomial time approximation algorithm whose ratio is equal to 5. Authors in [8] suggest a solution that uses artificial bee colony technique for placing RNs in the network. The RNs can be placed anywhere in the network without any restriction. The authors claimed that the proposed solution increases the network lifetime by selecting the best positions for the new added relays. However, the SNs can participate in the data forwarding, which reduces dramatically the network lifetime.

In [7], the authors consider the problem of constrained relay node placement, where the relay nodes can only be placed in a set of candidate positions. Previous work focused on the unconstrained version of the problem where relay nodes can be deployed anywhere. However, in practice there might be some physical constrains on the placement of relay nodes which makes the constrained version of the problem more realistic. In [7], the authors address the connectivity and the survivability problem in the context of constrained relay node placement. Approximation algorithms for solving these two problems are presented in [7] and their complexity is discussed. It is shown that these algorithms have a polynomial time complexity and a small approximation ratio. The obtained solutions for these problems have a single-tiered topology. The same topology has been considered in [3], but in an environment including energy harvesting enabled sensor nodes and non-harvesting (regular) nodes. The aim was to use only energy harvesting nodes to relay packet, with addition of a minimum number of relay nodes. A formulation based on minimum connected dominating set has been used to solve the problem.

Some works consider RNs placement with constraints related to specific applications, such as WSN deployment in tunnels [9], pipeline inspection [10], wind farm [11], and delay constraint applications [12]. In this work, we aim to tackle the problem with a general model that applies to a large set of applications. [13] considers deployment of fixed number of RNs and exploits convexity in special case of the network communication cost to determine optimal placement. [14] considers RNS placement jointly with sub-carrier allocation.

The authors of [15] study both the single-tiered and the two-tiered version of the relay node placement problem and provide an approximation algorithm for the single-tiered problem and a polynomial time approximation scheme for the two-tiered relay node placement problem. The survivability problem in single-tiered and two-tiered topology is investigated in [16]. The aim

in that work is to ensure that each sensor is connected to the BS through two disjoint paths. For the single-tiered problem a 16-approximation algorithm is provided, whereas for the two-tiered problem a $(20 + \epsilon)$-approximation algorithm is presented in [16], where ϵ is a positive constant. The connectivity problem in single-tiered and two-tiered topology is addressed in [17]. The relay node is assumed to have a wider communication range than the sensor node. The problems addressed in this paper are known to be NP-hard, thus, the authors derived heuristic algorithms to solve them. A 7-approximation algorithm for the single-tiered problem and a $(5+\epsilon)$ algorithm for the two-tiered problem are proposed in [17]. The problem of two-tiered constrained relay node placement is investigated in [1], while addressing both the connectivity and the survivability requirements. Most of the existing solutions for two-tiered problem are executed in three steps: $i)$ Connect sensor nodes to the base-stations; $ii)$ Connect remaining sensors to the relay nodes; $iii)$ Construct a connected Steiner tree to ensure the connection of all sensors to the base-stations. In contrast to these solutions, authors in [18] suggest to construct the two-tiered topology in a unified manner that gives a global view of the problem, and then intuitively helps to reduce the number of RNs required.

Usually, sensor nodes have limited resources in terms of energy power supply since they operate using batteries. In the single-tiered topology, it is highly probable that a sensor exhausts quickly its onboard energy supply and stops working. The wireless sensor network may be structurally damaged if many sensors run out of energy. Therefore, two-tiered topology [1], [15]–[17] would be more practical for limited resources WSN. For this reason, in this paper, we aim to build two-tiered topology to interconnect SNs with BSs. In contrast to existing solutions for constrained relay nodes placement, [1]–[3] and [7], this work is the sole solution that aims to enhance the network QoS taking into account the physical layer parameters when adding extra

relay nodes at a set of candidate positions.

III. PROBLEM FORMULATION AND NETWORK MODEL

A. Envisioned Architecture

The considered network comprises a set of SNs and a set of BSs referred to as $\mathcal{S}$ and $\mathcal{B}$, respectively. We denote by $\mathcal{Y}$ the set of possible locations where the RNs can be placed. The problem consists in finding the optimal deployment strategy for the RNs such that each sensor becomes connected to a BS. The relays ensure bidirectional communication between the SNs and the BSs by conveying the traffic generated at the SNs to the BSs and vice versa. The optimal choice of the RN positions should satisfy several objectives. Firstly, a good QoS should be ensured by offering reliable communication link between the SNs and the BSs. Secondly, we aim to interconnect the SNs to the RNs that have a good link in order to prevent the retransmissions, and thus extend the network lifetime. Thirdly, the number of deployed RNs should be minimized, to reduce the solution cost. In order to further increase the communication reliability, an automatic repeat request (ARQ) scheme is considered for forwarding the information. The ARQ scheme allows resending a packet until successful reception or a maximum number of retransmissions M is reached.

The BSs are connected together through wired links. Each SN in the network has a transmission power $P_{\mathcal{S}}$. The SNs have no harvesting capability and a limited battery capacity. Each RN have a transmission power $P_{\mathcal{R}}$, such that $P_{\mathcal{R}} \geq P_{\mathcal{S}}$. The RNs may have a harvesting capability, or they may be connected to the power grid. We consider a constrained relay node placement problem where the possible positions at which the RNs can be placed are limited. *Note that the number of possible deployment positions is larger than the number of relays that will be eventually introduced in the network.*

The notations used throughout this paper are summarized in Table I.

TABLE I

SUMMARY OF NOTATIONS.

Notation	Description
$P_{\mathcal{S}}$ and $P_{\mathcal{R}}$	The power of a SN and the power of a RN, respectively.
$\alpha_{u,v}$	The fading coefficient of the channel between the transmitting node u and the receiving node v.
$\alpha_{t,v}$	The fading coefficient of the channel between the interfering node t and the receiving node v.
$\gamma_{u,v}$	The instantaneous received signal-to-noise ratio at node v with mean value $\bar{\gamma}_{u,v}$.
$d_{u,v}$	The distance between node u and node v.
n_v	A zero-mean additive white Gaussian noise with variance N_0.
η	The path loss exponent.
$\mathsf{SINR}_{u,v}$	Signal-to-interference-plus-noise ratio (SINR) between nodes u and v.
γ_{th}	The receiver sensitivity threshold.
$\rho_{\mathcal{S}}$, $\rho_{\mathcal{R}}$, $\rho_{\mathcal{B}}$	The transmission ranges of SNs, RNs, and BSs, respectively.
$\mathcal{P}_{u,v}$	Outage probability for the link between nodes u and v.
$\mathcal{G} = (V, E)$	Graph $\mathcal{G}$ where V is the set of vertices and E is the set of edges.
$\mathbb{E}(T_r)$	The average number of retransmissions.
M	Maximum number of retransmissions for ARQ.
$\bar{R}$ and $\bar{P}$	Average transmission rate and average consumed power, respectively.
T_{soj} and W	The packet's sojourn time in the buffer and the average waiting time for a data packet, respectively.
η_{EE}	The energy efficiency.

B. Physical Model

The physical layer of the point-to-point communication between any couple of nodes in the network is described in this section. Let us denote by u the transmitting node, and by v the receiving node. The channel gain between these two nodes is referred to as $\alpha_{u,v}$. A Rayleigh

block-fading channel is considered, where the channel gain remains constant over one block[1] but changes independently from one block to another. In the proposed model, we take into consideration the interference from the other nodes in the environment on the receiving node v. The fading coefficient from a node t to the node v is referred to as $\alpha_{t,v}$, which follows a Rayleigh distribution. The received signal at a destination node v can be expressed as

$$y_v = \alpha_{u,v}\sqrt{P_u}x_u + \sum_{\substack{t \in \mathcal{N} \\ t \neq u,v}} \alpha_{t,v}\sqrt{P_t}x_t + n_v, \tag{1}$$

where P_u and P_t are the transmission powers at node u and at node t, respectively, and $\mathcal{N}$ is the set of all the transmitting nodes in the network. The symbols transmitted by node u and node t are referred to as x_u and x_t, respectively. The term n_v is a zero-mean additive white Gaussian noise with variance N_0. The value of N_0 is generally normalized to 1 [19]. We denote by, $\gamma_{u,v}$, the instantaneous received signal-to-noise ratio at node v given by $\gamma_{u,v} = P_u\alpha_{u,v}^2/N_0$ [19]. The mean value of $\gamma_{u,v}$ is denoted as $\bar{\gamma}_{u,v}$ which can be expressed as

$$\bar{\gamma}_{u,v} = \frac{P_u\mathbb{E}[\alpha_{u,v}^2]}{N_0}, \tag{2}$$

where $\mathbb{E}[\alpha_{u,v}^2]$ represents the channel variance and $\mathbb{E}[\cdot]$ denotes the expectation operator. Using a distance dependent path loss model, the channel variance can be determined as [20]

$$\mathbb{E}[\alpha_{u,v}^2] = \left(\frac{d_0}{d_{u,v}}\right)^{\eta}, \tag{3}$$

with $d_{u,v}$ referring to the distance between node u and node v and d_0 represents a reference distance typically set to 1 m [20]. The symbol η denotes the path loss exponent generally set to 2 [20].

---

[1]A block corresponds to the time duration necessary to send one packet.

In our physical model, the effects of both path loss and fast fading are taken into account, while the impact of shadowing on system performance is neglected. This assumption leads to a simplified model and makes the analysis tractable. Note also that the effect of shadowing can be neglected if the SNs are deployed in large fields with no obstruction between the transmitter and the receiver. The instantaneous received SINR at node v is defined as [19]

$$\mathsf{SINR}_{u,v} = \frac{P_u \alpha_{u,v}^2}{N_0 + \sum_{t \in \mathcal{N}}^{t \neq u,v} P_t \alpha_{t,v}^2} = \frac{\gamma_{u,v}}{1 + \sum_{t \in \mathcal{N}}^{t \neq u,v} \gamma_{t,v}}. \tag{4}$$

C. Network Model

As mentioned in Section III-A, we are interested to place the minimum number of RNs, say the set $\mathcal{R}$, in candidate positions[2] $\mathcal{Y}$ to offer bidirectional communication between the SNs of $\mathcal{S}$ and the BSs, $\mathcal{B}$, while ensuring a high QoS and reducing the energy consumption. RNs can be viewed as energy unconstrained nodes that sustain to uninterrupted operation. BSs can communicate directly via either the wired networks or the satellites [1], and they can be viewed as a clique. It is also assumed that the BSs have a larger transmit power than the RNs, and the latter have a larger transmit power than the SNs ($P_\mathcal{R} \geq P_\mathcal{S}$). In this paper, selection of high quality links is envisaged to sustain the traffic forwarded from and toward the BSs. The use of high quality links by the RNs reduces the packet retransmissions, which has a positive impact on the data transfer delay and the network goodput. The use of high quality links by the SNs not only reduces the data transfer delay and increases the network goodput, but also has a positive impact on the energy consumption at these nodes, which leads to increase the network lifetime.

To reduce the complexity of the proposed solution, the search space of candidate parents [3]

---

[2]The relay nodes can only be placed in predefined locations which are called throughout the paper candidate positions.

[3]The candidate parents are the relay nodes to which a sensor node can be connected. In contrast, a parent is the relay node to which the sensor is actually connected and which is used to forward the data traffic from the sensor.

is reduced for each node u, to its feasible set. It is intuitive that if two nodes are not able to communicate in an interference-free environment, they will definitely fail to communicate in presence of interference. As the transmission power is a good indicator for communication possibility in interference-free environment, it is used to limit the search space for candidate parents. Therefore, we define the transmission range ρ_u for each node u in the network according to its transmission power, and then the set of potential parents of a node u is restricted to the nodes that are within a distance ρ_u from node u. Indeed, a node u can select node v as its parent iff the euclidian distance between u and v does not exceed ρ_u (i.e., $d(u,v) \leq \rho_u$). The following lemma specifies how to determine the transmission range, ρ_u, of a node u according to its power, P_u.

Lemma 1. *Node u can select v as a parent node iff v is within a distance less than ρ_u from node u, where ρ_u is defined as follows:*

$$\rho_u = d_0 \sqrt[\eta]{\frac{P_u}{N_0 \gamma_{\text{th}}}}, \tag{5}$$

where P_u is the transmitter power, γ_{th} is the receiver sensitivity threshold, and η is the path loss exponent.

Proof: Using (2) and (3), the average received signal-to-noise ratio at node v can be expressed as

$$\bar{\gamma}_{u,v} = \frac{P_u \mathbb{E}[\alpha_{u,v}^2]}{N_0} = \frac{P_u}{N_0} \left(\frac{d_0}{d_{u,v}}\right)^\eta.$$

In average, the transmission from u to v succeeds only if the average received signal-to-noise ratio at v exceeds the sensitivity, γ_{th}, of the receiver v, i.e., $\bar{\gamma}_{u,v} \geq \gamma_{\text{th}}$. Let ρ_u denote the

maximum range within which the node u succeeds to transmit to node v. It follows that

$$\gamma_{\text{th}} = \frac{P_u}{N_0} \left(\frac{d_0}{\rho_u} \right)^{\eta}, \qquad \text{thus} \qquad \rho_u = d_0 \sqrt[\eta]{\frac{P_u}{N_0 \gamma_{\text{th}}}}.$$

■

From Lemma 1, we can deduce the transmission ranges $\rho_{\mathcal{S}}$, $\rho_{\mathcal{R}}$, and $\rho_{\mathcal{B}}$, respectively of SNs, RNs, and BSs, from their transmission power. From Lemma 1, it is obvious that $\rho_{\mathcal{B}} \geq \rho_{\mathcal{R}}$ and $\rho_{\mathcal{R}} \geq \rho_{\mathcal{S}}$. In what follows, the undirected graph $G = (V, E)$ will be defined. This graph will help to reduce the complexity of the proposed solution.

Definition 1. *The communication between nodes is modeled as a graph $G = (V, E)$, where V is the set of all nodes in the network augmented with all candidate positions of RNs, i.e., $V = \mathcal{B} \cup \mathcal{S} \cup \mathcal{Y}$, and $E = E_{\mathcal{SS}} \cup E_{\mathcal{SB}} \cup E_{\mathcal{SY}} \cup E_{\mathcal{YY}} \cup E_{\mathcal{YB}} \cup E_{\mathcal{BB}}$ is the set of edges in G. $E_{\mathcal{ZW}}$ denotes the edges from the vertices of $\mathcal{Z}$ to the vertices of $\mathcal{W}$. An edge $(u, v) \in E$ iff the two nodes u and v are within the transmission range of each other. The edges E are formed as follow: For any two BSs $b_i, b_j \in \mathcal{B}$, an edge (b_i, b_j) is constructed. An edge (u, v) is also formed between a SN, $u \in \mathcal{S}$, and every vertex $v \in V$ iff $d_{u,v} \leq \rho_{\mathcal{S}}$. For any vertex $u \in \mathcal{Y}$, and $v \in \mathcal{Y} \cup \mathcal{B}$, an edge (u, v) is constructed iff $d_{u,v} \leq \rho_{\mathcal{R}}$. It is guaranteed that the formed graph G ensures the bidirectional communication between the nodes in the network.*

Theorem 1. *$\forall u \in V$ and $\forall v \in V$, vertex v fails to receive correctly a packet transmitted from vertex u if $\text{SINR}_{u,v}$ falls below a threshold γ_{th}, i.e. $\text{SINR}_{u,v} < \gamma_{\text{th}}$. This event is known as an outage event and occurs with a probability, $\mathcal{P}_{u,v}$, which can be expressed as*

$$\mathcal{P}_{u,v} = 1 - \sum_{t \in \mathcal{N}}^{t \neq u,v} C_{t,v} \frac{\bar{\gamma}_{u,v}}{\bar{\gamma}_{u,v} + \bar{\gamma}_{t,v} \gamma_{\text{th}}} \exp \left(-\frac{\gamma_{\text{th}}}{\bar{\gamma}_{u,v}} \right), \tag{6}$$

where $C_{t,v} = \prod_{z \in \mathcal{N}}^{z \neq u,v,t} \frac{\bar{\gamma}_{t,v}}{\bar{\gamma}_{t,v} - \bar{\gamma}_{z,v}}$ and $\mathcal{N}$ is the set of transmitting nodes.

Proof: See Appendix A.   ■

From (6), we can conclude that the outage probability $\mathcal{P}_{u,v}$ increases proportionally with the receiver sensitivity threshold γ_{th}. In fact, the second term in the right hand side of (6) decreases with γ_{th}. For large values of γ_{th}, the term $\sum_{t \in \mathcal{N}}^{t \neq u,v} C_{t,v} \frac{\bar{\gamma}_{u,v}}{\bar{\gamma}_{u,v} + \bar{\gamma}_{t,v} \gamma_{\text{th}}} \exp\left(-\frac{\gamma_{\text{th}}}{\bar{\gamma}_{u,v}}\right) \to 0$, and thus the outage probability $\mathcal{P}_{u,v} \to 1$. The receiver sensitivity threshold γ_{th} indicates the threshold that the SINR should exceed so that the packet is successfully received. As the sensitivity threshold of the receiver increases, its ability to detect weak signals decreases. The larger the receiver sensitivity threshold is, the lower the probability of successful packet reception becomes.

Definition 2. *To capture the effect of the outage probability on the network performance, the communication between nodes is modeled by a weighted graph $\mathcal{G} = (V, E, \omega)$, where V is the set of vertices and E is the set of weighted edges. ω represents the weight of these edges, $\omega_{u,v} \in \omega$, which is formed as follows:*

$$\forall (u,v) \in E : \omega_{u,v} = \frac{1}{2}\left(\mathcal{P}_{u,v} + \mathcal{P}_{v,u}\right). \tag{7}$$

As stated in Definition 2, the weight associated to the edge (u, v) is the mean value of the outage probabilities for the links (u, v) and (v, u). This is to reflect the quality of the bidirectional communication link between the vertices. The choice of the weight value according to (7) fulfills this requirement. By minimizing the weights $\omega_{u,v}$, the outage probabilities $\mathcal{P}_{u,v}$ and $\mathcal{P}_{v,u}$ are also minimized.

IV. Realistic constrained Relay Nodes Placement (RRPL) Algorithm

As pointed above, the selection of links with high quality to sustain the traffic forwarded from and toward the BSs has a positive impact on the network lifetime and the QoS in the network. In this section, we present our *RRPL* algorithm for constrained relay nodes placement in WSN,

that has a polynomial runtime complexity. The idea behind *RRPL* is to use a polynomial time α-approximation algorithm of the Steiner tree problem ($\mathcal{ASTP}$) to interconnect the SNs with the BSs. The Steiner tree problem in our case is defined as follows: Given a set $\mathcal{S} \cup \mathcal{B}$ of nodes in not connected graph G, determine a set $\mathcal{Y}$ of Steiner points that should be added to the graph such that the minimum spanning tree (MST) cost over $\mathcal{S} \cup \mathcal{B} \cup \mathcal{Y}$ is minimized. The use of $\mathcal{ASTP}$ allows us to use the minimum number of relay nodes, as well as it minimizes the outage probability between nodes. In the sequel of this section, the *RRPL* algorithm will be presented, then its runtime complexity will be analyzed.

A. Algorithm Description

In the balance of this section, we use Algorithm 1 to explain the functionality of *RRPL*. Fig. 1 serves as a detailed example that will be referenced to illustrate the operations of *RRPL*. In this example, two BSs ($\mathcal{B}$) are considered, they are numbered 0 and 1, respectively. In Fig. 1(a), the circles and the squares represent the SNs ($\mathcal{S}$) and candidate relay positions ($\mathcal{Y}$), respectively. Before starting the execution of *RRPL*, we have to deduce the communication range of nodes ($\rho_{\mathcal{B}}$, $\rho_{\mathcal{S}}$ and $\rho_{\mathcal{R}}$) through Lemma 1 and using the positions of $\mathcal{B} \cup \mathcal{S} \cup \mathcal{Y}$, receiver sensitivity threshold, and transmission powers of SNs ($P_{\mathcal{S}}$) and RNs ($P_{\mathcal{R}}$). Then $G = (V, E)$ would be constructed using Definition 1, as depicted in Fig. 1(a). In this figure, a dashed edge between vertices in G indicates that each vertex can select the other as a parent. Fig. 1(b) shows the outage probability of different edges in the network. Fig. 1(c) depicts the weighted graph $\mathcal{G} = (V, E, \omega)$, which is constructed from G using Definition 2. A weight 0 is assigned to edges between BSs in order to favor the selection of the links between them when forming the Steiner tree in $\mathcal{G}$. Algorithm 1 uses $\mathcal{G}$, $\mathcal{S}$, $\mathcal{Y}$, $\mathcal{B}$ and $\mathcal{ASTP}$ as inputs, while it has as output the set of activated positions $\mathcal{R}$ where the RNs should be placed. We assume that the BSs of $\mathcal{B}$ and the RNs of $\mathcal{Y}$ belong to the
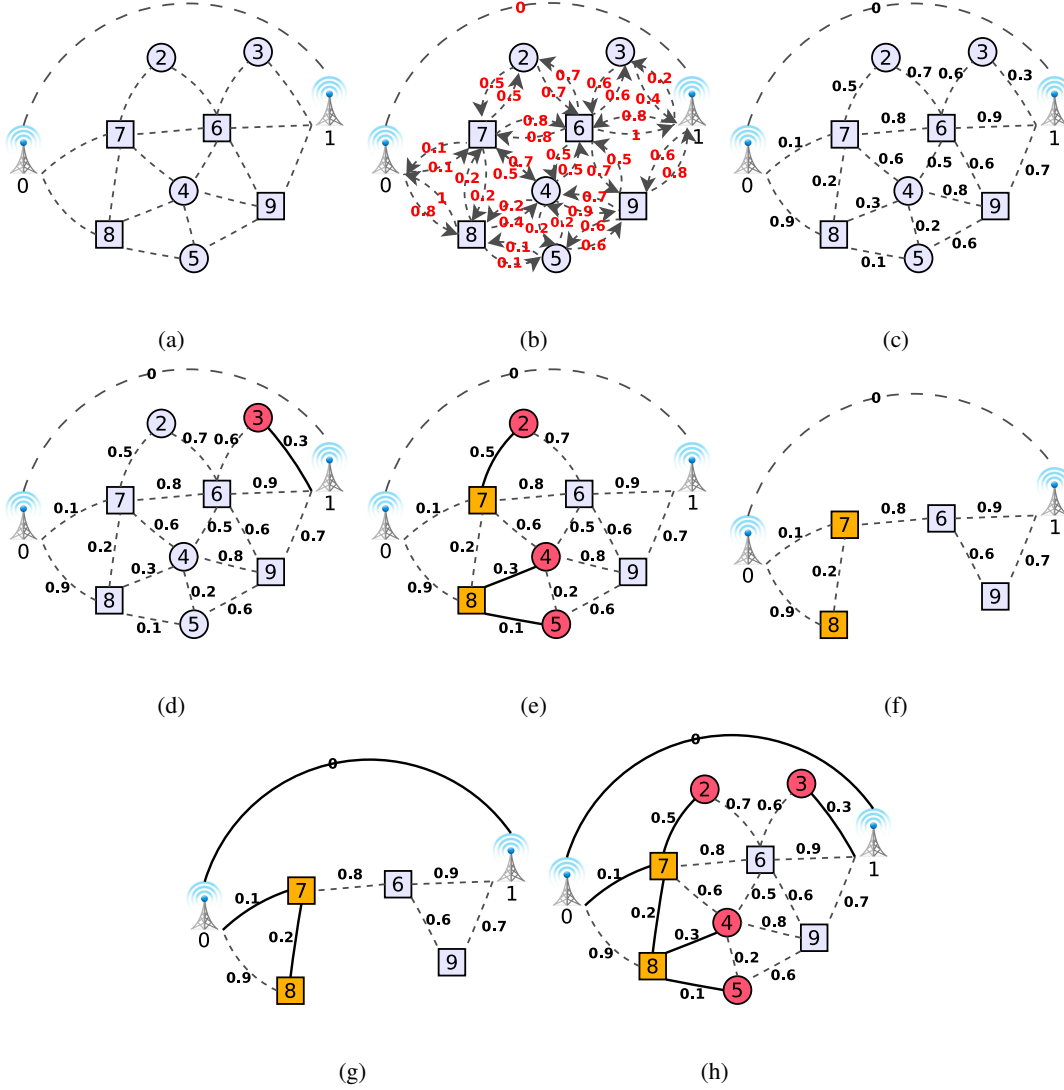
Fig. 1. Illustrative example that shows the execution of *RRPL*

same connected component $\mathcal{C}$ in $\mathcal{G}$, as well as $\mathcal{B} \cup \mathcal{Y}$ is a vertex cover of $\mathcal{G}$. Otherwise, *RRPL* cannot interconnect $\mathcal{S}$ to $\mathcal{B}$, using a subset of vertices from $\mathcal{Y}$. Actually, in this case the given topology does not have a feasible solution.

RRPL forms the activated positions $\mathcal{R}$ by executing three main steps: (i) Interconnecting SNs to the BSs (Algorithm 1: lines $1 - 13$); (ii) Interconnecting SNs to RNs of $\mathcal{Y}$ (Algorithm 1: lines $14 - 34$); (iii) Interconnecting the activated RNs to the BS via $\mathcal{ASTP}$ (Algorithm 1: lines

$35 - 38$). Steps (i) and (ii) are executed before forming the Steiner tree in order to ensure the creation of two-tiered topology. As we will see later in this section, after the execution of steps (i) and (ii), the SNs will be removed from the graph that will be used as input to $\mathcal{ASTP}$. This means that the Steiner tree would be formed only through $\mathcal{Y}$ without using any SNs. In the end of steps (i), (ii) and (iii), SNs will connected directly to BSs or through a set of RNs from $\mathcal{Y}$, which is a two-tiered topology.

1) Interconnecting $\mathcal{S}$ to $\mathcal{B}$: Firstly, we initialize the variable $Parent$ that represents the relationship between the nodes and their parents (Algorithm 1: line 1). Then, *RRPL* removes the SNs that are neighbors to the BSs. Assigning the BSs as parents for those sensors reduces the path length, and thus decreases the data transfer delay. For the SN that has multiple BS neighbors, the BS that has a good link with that sensor will be selected as parent in order to reduce the energy consumption and increase the QoS (Algorithm 1: lines $4 - 9$). To ensure the construction of two-tiered topology, those SNs are removed from $\mathcal{G}$ (Algorithm 1: lines $10 - 12$). As depicted in Fig. 1(d), *RRPL* assigns BS 1 to SN 3 as parent, and then removes SN 3 from $\mathcal{G}$.

2) Interconnecting $\mathcal{S}$ to $\mathcal{Y}$: In this step, the remaining SNs $\mathcal{S}$ would be attached to the candidate relay positions $\mathcal{Y}$ (Algorithm 1: lines $14 - 34$). This is in order to prevent the selection of SNs as parents, when forming the Steiner tree, and hence ensuring the construction of two-tiered topology. These candidate relay positions would be activated and considered with the BSs as source nodes for the Steiner tree algorithm. *RRPL* starts by initializing the source nodes of Steiner tree with $\emptyset$ (Algorithm 1: line 14). Then, if there are still some SNs in $\mathcal{G}$ (Algorithm 1: line 15), *RRPL* performs the following tasks: (i) activate the relay position that leads to reduce the cost and enhance the QoS; (ii) select the activated relay as parent for its neighbor SNs and

---

Algorithm 1 Realistic Relay node PLacement (RRPL) Algorithm

Input:
$\mathcal{G} = (V, E, \omega)$: Weighted communication graph, where $E = E_{SS} \cup E_{S\mathcal{Y}} \cup E_{S\mathcal{B}} \cup E_{\mathcal{Y}\mathcal{Y}} \cup E_{\mathcal{Y}\mathcal{B}}$

$\mathcal{B}$: Set of base stations.

$\mathcal{S}$: Set of sensor nodes in the network.

$\mathcal{Y}$: Set of candidate relay nodes in the network.

$\mathcal{AST\,P}(\mathcal{G}, \omega, S)$: An approximation algorithm to obtain a Steiner tree from the weighted gra
  where S are the source nodes.

Output:
$\mathcal{R}$: The set of relay nodes positions that should be activated.

1: $Parent = \{\}$;
2: **for all** $sn \in \mathcal{S}$ **do**
3:    $\omega_{min} = \infty$;
4:    **for all** $(sn, bst) \in E_{S\mathcal{B}}$ **do**
5:       **if** $\omega_{min} > \omega_{sn,bst}$ **then**
6:          $\omega_{min} = \omega_{sn,bst}$;
7:          $Parent[sn] = bst$;
8:       **end if**
9:    **end for**
10:   **if** $\omega_{min} \neq \infty$ **then**
11:      $V = V \setminus \{sn\}$;
12:   **end if**
13: **end for**
14: $Src = \emptyset$;
15: **while** $\mathcal{S} \cap V \neq \emptyset$ **do**
16:   $\mathcal{A} = \infty$;
17:   $snN = \emptyset$;
18:   $nb\omega = \infty$;
19:   **for all** $rly \in \mathcal{Y} \setminus Src$ **do**
20:      $snNrly = N(\mathcal{G}, rly) \cap \mathcal{S}$; // $N(\mathcal{G}, rly)$ *is a function that returns the neighbors of rly in* $\mathcal{G}$.
21:      **if** $(size(snNrly) > size(snN) \vee (size(snNrly) = size(snN) \wedge nb\omega > \sum_{n \in snNrly} \omega_{n,rly})$ **then**
22:         $\mathcal{A} = rly$;
23:         $snN = snNrly$;
24:         $nb\omega = \sum_{n \in snNrly} \omega_{n,rly}$;
25:      **end if**
26:   **end for**
27:   **if** $\mathcal{A} \neq \infty$ **then**
28:      $Src = Src \cup \{\mathcal{A}\}$;
29:      **for all** $sn \in snN$ **do**
30:         $V = V \setminus \{sn\}$;
31:         $Parent[sn] = \mathcal{A}$;
32:      **end for**
33:   **end if**
34: **end while**
35: $Src = Src \cup \mathcal{B}$;
36: $(\mathcal{V}, \mathcal{E}) = \mathcal{AST\,P}(\mathcal{G}, \omega, Src)$;
37: $\mathcal{R} = \mathcal{V} \cap \mathcal{Y}$;
38: **Return** $\mathcal{R}$;

---

remove the latter from $\mathcal{G}$.

Firstly, *RRPL* activates the relay position $\mathcal{A}$ that has the largest number of neighboring sensors in order to reduce as much as possible the number of RNs that should be placed in the network and thus decrease the solution cost. If there is more than one RN which fulfils this condition, *RRPL* activates the relay position that reduces the energy consumption and data transfer delay (Algorithm 1: lines $16 - 26$). As depicted in Fig. $1(e)$, activating relay position 9 would have a negative impact on the energy consumption and data transfer delay at sensors 4 and 5. For this reason, *RRPL* activates the relay position 8. In a similar way, activating the relay position 6 to

handle the traffic of SN 2 has a negative impact on the network performances. Secondly, the activated relay $\mathcal{A}$ is added to the sources of the Steiner tree (Algorithm 1: line 28). Afterwords, the SNs that are neighbors to $\mathcal{A}$ select the latter as parent and then these SNs are removed from $\mathcal{G}$ (Algorithm 1: lines $29 - 32$). As depicted in Fig. $1(f)$, the SNs that are neighbors to the activated relay positions, 7 and 8, are removed from $\mathcal{G}$.

3) Interconnecting $\mathcal{R}$ to $\mathcal{B}$: Firstly, the sources of the Steiner tree, formed in the previous step, would be augmented with the BSs $\mathcal{B}$ (Algorithm 1: line 35). Then, *RRPL* applies the algorithm [21] to achieve a minimum weighted Steiner tree for interconnecting the sources in $\mathcal{G}$ (Algorithm 1: line 36). The candidate positions of RNs in the Steiner tree would be identified and then activated (Algorithm 1: lines $36 - 37$). Fig. $1(g)$ shows the minimum weighted Steiner tree, which comprises the BSs and the candidate RNs positions 7 and 8. In this figure, instead of choosing the BS 0 as parent for vertex 8, vertex 7 is selected as parent that leads to interconnect vertex 8 with the lowest outage probability to $\mathcal{B}$, which has a positive impact on the network QoS. Fig. $1(h)$ shows the final tree that is constructed by *RRPL* algorithm.

Note that the initial set of candidate positions comprises the vertices $6, 7, 8$ and 9. Among these candidate RN positions, only position 7 and 8 are activated by *RRPL*. In other words, we have to deploy 2 RNs in position 7 and 8, whereas no RNs will be deployed in positions 6 and 9. This choice ensures the required QoS by adding only two RNs to the network.

B. Time Complexity Analysis of RRPL

Theorem 2. *The runtime of Algorithm 1 is asymptotically bounded by $O(|E_{\mathcal{SB}}| + |\mathcal{S}|(|E_{\mathcal{SY}}| + |\mathcal{Y}|) + T_{\mathcal{ASTP}})$, where $T_{\mathcal{ASTP}}$ is the time for executing the $\mathcal{ASTP}$.*

Proof: Starting with the first loop, from line 2 through 13, operations of line 2 and 3 repeat

$|\mathcal{S}|$ times, while the inner loop repeats $|E_{\mathcal{SB}}|$ times in total. The complexity of the first loop (from line 2 to 13) is then $O(|\mathcal{S}| + |E_{\mathcal{SB}}|)$. Now let us analyze the $while$ loop starting at line 15 that includes two inner loops. All instruction of the first inner loop from line 19 to 26 may be considered as atomic operations, except the summation $\sum_{n \in SnRly} \omega_{n,rly}$. The number of steps to calculate the latter is $|sn\mathcal{N}rly|$ in every pace of the for loop, which results in $|E_{\mathcal{SY}}|$ steps in total at the exit of the for loop, and thus $|E_{\mathcal{SY}}||\mathcal{S}|$ steps in total. For the other operations, the number of steps is simply $|\mathcal{Y}||\mathcal{S}|$. Every pace in the for loop of line 29 repeats $|sn\mathcal{N}|$ times (i.e., the number of sensor nodes adjacent to the selected relay node). As a relay node cannot be selected more than once (it is removed from $\mathcal{Y}$ in line 19), the total repetitions of the for loop operations in the $while$ loop can be bounded by $|E_{\mathcal{SY}}|$. The last step of the algorithm has a complexity of $T_{\mathcal{ASTP}}$. Consequently, the total complexity can be obtained by summing up all the previous expressions: $O(|\mathcal{S}| + |E_{\mathcal{SB}}| + |E_{\mathcal{SY}}||\mathcal{S}| + |\mathcal{Y}||\mathcal{S}| + |E_{\mathcal{SY}}| + T_{\mathcal{ASTP}})$.

As the terms $|\mathcal{S}|$ and $|E_{\mathcal{SY}}|$ are dominated by $|E_{\mathcal{SY}}||\mathcal{S}|$, the following complexity is obtained by eliminating them: $O(|E_{\mathcal{SB}}| + |\mathcal{S}|(|E_{\mathcal{SY}}| + |\mathcal{Y}|) + T_{\mathcal{ASTP}})$. ∎

It is known that Steiner tree heuristics have polynomial time complexity, so the previous expression is also polynomial.

V. THEORETICAL ANALYSIS

In order to improve the link reliability between two nodes u and v, we use an ARQ scheme, where the receiving node sends an ACK message to the transmitting node to indicate successful reception. If the data packet is received correctly, the transmitting node sends the next data packet. Otherwise, the same packet is resent. For a given data packet, the transmission terminates at the successful reception, or after the maximum number of retransmissions M is reached. The number of ARQ retransmissions required for successful packet receptions varies according to the channel

conditions and the interference level. If the channel conditions are good and the interference level is low, a single transmission could be sufficient for successful packet reception. In the case of bad channel conditions and high interference levels, M retransmissions might be required to successfully receive one data packet. It might be possible that even after M retransmissions, the packet is not successfully received. In this case, the packet is discarded. The use of a finite value of M allows limiting the delay and avoiding the congestion in the network.

A. Average Number of Retransmissions

If a packet reception fails, the transmitting node keeps sending that packet until successful reception or a maximum number of retransmissions M is reached. A reception failure occurs if the SINR level is below the sensitivity threshold of the receiving node. The SINR level varies randomly from one transmission to another. Thus, the number of retransmissions T_r needed to deliver successfully a packet is a random variable which varies depending on the channel conditions. The expression of the average number of retransmissions $\mathbb{E}(T_r)$ with a maximum number of retransmissions M is derived in Appendix B as

$$\mathbb{E}(T_r) = 1 + \sum_{m=1}^{M-1} P(F^1, ..., F^m), \tag{8}$$

where F^m denotes the event reception failure at the mth retransmission. The notation $P(F^1, ..., F^m)$ represents the probability of a reception failure at the 1st, 2nd, ..., mth retransmissions. Such an event occurs if the SINR at all these retransmissions is below the sensitivity threshold of the receiver. To show that the failure events $F^1, F^2, ..., F^m$ are independent, it suffices to show that the SINRs at the 1st, 2nd, ..., mth retransmissions are independent. In the following, we provide arguments supporting the independency of the SINRs at the 1st, 2nd, ..., mth retransmissions.

Using the expression of the SINR in (4), it is clear that $\mathsf{SINR}_{u,v}$ depends only on the signal-to-noise ratios $\gamma_{u,v}$ and $\gamma_{t,v}$. In our physical model, we assume a block-fading channel for which the channel coefficients vary independently from one transmission to another. This implies that the signal-to-noise ratios $\gamma_{u,v}$ and $\gamma_{t,v}$ vary independently for different transmissions. More specifically for any two retransmissions i and j, we have two realizations of the random variable $\gamma_{u,v}$ denoted $\gamma_{u,v,i}$ and $\gamma_{u,v,j}$. These realizations $\gamma_{u,v,i}$ and $\gamma_{u,v,j}$ are independent due to the fact that the fading channel is a block-fading channel. Moreover, the realizations $\gamma_{t,v,i}$, $\gamma_{t,v,j}$, $\gamma_{u,v,i}$ and $\gamma_{u,v,j}$ are all independent. It follows that the SINRs at different retransmissions are independent. Consequently, the failure events $F^1, F^2, ..., F^m$ are independent and we can write

$$P(F^1, ..., F^m) = P(F^1) \times \cdots \times P(F^m). \tag{9}$$

In the following, we derive an expression of the probability $P(F^m)$ of a reception failure at the mth retransmission. A failure in the reception occurs at the mth retransmission if the SINR at this step falls below a certain threshold value γ_{th}. Thus, we can determine $P(F^m)$ as

$$P(F^m) = P(\mathsf{SINR}_{u,v,m} \leq \gamma_{th}) = F(\gamma_{th}), \tag{10}$$

where $F(\cdot)$ is the CDF of the SINR between nodes u and v at the mth retransmission denoted as $\mathsf{SINR}_{u,v,m}$.

In our channel model, we consider a block-fading channel where the channel remains constant over one retransmission but changes independently from one retransmission to another. At each retransmission the channel realization is an output of a random generator. Note that these channel realizations are independent and identically distributed. In other words, the SINRs at different retransmissions are independent and have the same distributions and thus the same CDF. It follows that the probability of a reception failure is the same at any given retransmission $m(m =$

$1, \ldots, M$). Consequently,

$$P(F^m) = F(\gamma_{th}) = \mathcal{P}_{u,v} \quad \text{for} \quad m = 1, \ldots, M \tag{11}$$

$$P(F^1, ..., F^m) = P(F^1) \times \cdots \times P(F^m) = (F(\gamma_{th}))^m = (\mathcal{P}_{u,v})^m, \tag{12}$$

where $\mathcal{P}_{u,v}$ is the outage probability whose expression is provided in Theorem 1.

Using (8) and (12), we can determine the average number of retransmissions as

$$\mathbb{E}(T_r) = 1 + \sum_{m=1}^{M-1} P(F^1, ..., F^m) = 1 + \sum_{m=1}^{M-1} (\mathcal{P}_{u,v})^m \tag{13}$$

$$= \sum_{m=0}^{M-1} (\mathcal{P}_{u,v})^m = \frac{1 - (\mathcal{P}_{u,v})^M}{1 - \mathcal{P}_{u,v}}. \tag{14}$$

In (14), we used the identity for the sum of a geometric progression [22, Eq. (0.112)]. Using (13), we can conclude that the average number of retransmissions $\mathbb{E}(T_r)$ increases as the outage probability increases. In Section III-C, we showed that the outage probability $\mathcal{P}_{u,v}$ increases proportionally with the receiver sensitivity threshold γ_{th}. Consequently, the average number of retransmissions $\mathbb{E}(T_r)$ increases with the receiver sensitivity threshold γ_{th}. Moreover, it can be deduced from (13) that if the value of M decreases, the average number of retransmissions $\mathbb{E}(T_r)$ declines.

B. Average Transmission Rate

For each packet, the number of retransmissions varies from one data packet to another depending on the channel conditions and the interference level. Thus, the transmission rate R is a random variable. In this section, we derive an expression for the average transmission rate $\bar{R}$. Towards this end, let us assume that the transmitted data packet contains b information bits. In each retransmission, L symbols are sent to the receiving node. The rate of the first ARQ transmission in bits per channel use can be written as $R_1 = b/L$.

Note that the rate varies depending on the number of retransmissions used to send a data packet. For instance, if a packet requires m retransmissions to be successfully received, then for this packet the b information bits are sent over $L \times m$ symbols. In this case, the transmission rate $R = b/(L \times m) = R_1/m$. In average, b information bits are sent over $L \times \mathbb{E}(T_r)$ and thus the average transmission rate $\bar{R}$ can be expressed as

$$\bar{R} = \frac{b}{L \times \mathbb{E}(T_r)} = \frac{R_1}{\mathbb{E}(T_r)}. \tag{15}$$

The expression in (15) is valid for the case when $M \to \infty$. However, for a finite value of M and in case of bad channel conditions or high level of interference, the packet transmission may still fail even after the M retransmissions. In such case, the average rate would be equal to zero. This event occurs with a probability $P(F^1, ..., F^M) = (\mathcal{P}_{u,v})^M$. Consequently, using (15), we can determine the average transmission rate for the case of finite value of M as

$$\bar{R} = \frac{R_1(1 - P(F^1, ..., F^M))}{\mathbb{E}(T_r)} = R_1 \left(1 - \mathcal{P}_{u,v}\right). \tag{16}$$

The average transmission rate $\bar{R}$ decreases as the outage probability $\mathcal{P}_{u,v}$ increases. Since the outage probability increases as the threshold γ_{th} increases (as shown in Section III-C), thus, the average transmission rate decreases as the receiver sensitivity threshold γ_{th} increases.

C. Delay Analysis

In this section, we focus on the delay analysis. We consider that the nodes in the network are equipped with a buffer to store the packets before transmission. This will allow controlling the packet flow in the network and reducing network congestion. We assume Poisson arriving packets at the buffer with arrival rate λ. We analyse in this section the average sojourn time in the buffer and the average waiting time for a data packet. The average sojourn time in the

buffer is the average time elapsed from the arrival of the packet to a node u until its successful reception at the next node in the tree. The average waiting time for a data packet is the time spent by the packet in the buffer of a node u which begins from the arrival of that packet at the buffer and the start of its transmission.

The packet's sojourn time in the buffer can be evaluated using the Pollaczek-Khinchin equation as [23]

$$T_{\text{soj}} = W + \mathbb{E}(T_r)T_F, \tag{17}$$

where T_F is the frame duration and W is the average waiting time for a data packet. The expression of the average number of transmissions $\mathbb{E}(T_r)$ is provided by (14).

The average waiting time for a data packet can be determined as [23]

$$W = \frac{\lambda \mathbb{E}(T_r^2)T_F^2}{2(1 - \rho)} + \frac{T_F}{2}, \tag{18}$$

where ρ is a parameter that should satisfy the following stability condition

$$\rho = \lambda \mathbb{E}(T_r)T_F < 1. \tag{19}$$

The term $\mathbb{E}(T_r^2)$ stands for the second-order moment of the number of retransmissions and can be expressed as [24]

$$\mathbb{E}(T_r^2) = 1 + \sum_{m=1}^{M-1}(2m+1)P(F^1, ..., F^m). \tag{20}$$

Using [22, Eq. (0.113)] and (12), we can further simplify the expression of $\mathbb{E}(T_r^2)$ as

$$\mathbb{E}(T_r^2) = 1 + \sum_{m=1}^{M-1}(2m+1)\left(\mathcal{P}_{u,v}\right)^m = \sum_{m=0}^{M-1}(2m+1)\left(\mathcal{P}_{u,v}\right)^m$$
$$= \frac{1 - (2M-1)\left(\mathcal{P}_{u,v}\right)^M}{1 - \mathcal{P}_{u,v}} + \frac{2\mathcal{P}_{u,v}\left(1 - \left(\mathcal{P}_{u,v}\right)^{M-1}\right)}{\left(1 - \mathcal{P}_{u,v}\right)^2}. \tag{21}$$

From (21), we can conclude that the second-moment $\mathbb{E}(T_r^2)$ of the number of retransmissions increases as the outage probability increases. Using (18), it can be noticed that the average

waiting time W is proportional to the second-moment $\mathbb{E}(T_r^2)$. Similarly, from (17), we can deduce that the sojourn time T_{soj} is proportional to both the average and the second-moment of the number of retransmissions $\mathbb{E}(T_r)$ and $\mathbb{E}(T_r^2)$, respectively. It follows that the sojourn time and the average waiting time increase as the outage probability increases. On the other hand, since the outage probability increases with the receiver sensitivity threshold γ_{th} (as discussed in Section III-C), then the sojourn time and the average waiting time increase proportionally with the receiver sensitivity threshold.

D. Energy Efficiency

In this section, we study the energy efficiency of the ARQ scheme. The energy efficiency, denoted by η_{EE}, is a powerful tool for evaluating the efficiency of energy consumption in a communication system. The energy efficiency indicates how much energy is spent in average for each successfully received bit. The energy efficiency is defined as the ratio of the throughput and the average consumed power and is given by [25]

$$\eta_{EE} = \frac{\bar{R}}{\bar{P}} = \frac{R_1 \left(1 - (\mathcal{P}_{u,v})^M \right)}{\bar{P}}, \tag{22}$$

where the quantity $\bar{P}$ refers to the average consumed power. The expression of $\bar{R}$ is given by (16). The term $\bar{P}$ can be determined for the ARQ scheme as the product of two terms: the power per retransmission, $P_u{}^4$, at a given node u and the average number of transmission $\mathbb{E}(T_r)$. It follows that the average consumed power $\bar{P}$ can be expressed as

$$\bar{P} = P_u \cdot \mathbb{E}(T_r) = P_u \cdot \frac{1 - (\mathcal{P}_{u,v})^M}{1 - \mathcal{P}_{u,v}}. \tag{23}$$

---

[4]We assume that the power is constant for all retransmissions.

As mentioned in Section III-C, the receiver sensitivity threshold γ_{th} has a negative impact on the average number of retransmissions $\mathbb{E}(T_r)$. From (23), it can be deduced that the increase in $\mathbb{E}(T_r)$ leads to a larger average transmission power $\bar{P}$, and thus to a shorter network lifetime. While the increase of M has a negative impact on $\bar{P}$. Actually, as M increases more retransmissions occur (as discussed in Section V-A) which results in a higher consumed power.

Using (22) and (23), we can simplify the expression for the energy efficiency as

$$\eta_{EE} = \frac{R_1 \left(1 - \mathcal{P}_{u,v}\right)}{P_u}. \tag{24}$$

VI. SIMULATION

The proposed algorithm *RRPL* is evaluated through simulation and compared against three competitive solutions *TTCR* [1], *RNPEC* [7] and *ACRN* [2]. *RRPL* is compared to these algorithms and evaluated in terms of the following metrics: i) The end-to-end delay, defined as the time that a packet spends to travel from $\mathcal{S}$ to $\mathcal{B}$; ii) The goodput, defined as the number of useful information bits reaching the BSs $\mathcal{B}$; iii) The network lifetime, defined as the average time to battery drain out of SNs; iv) The cost, defined as the number of RNs $\mathcal{R}$ that should be added to the network. In the simulation results, each plotted point represents the average of 60 executions. The plots are presented with 95% confidence interval. In our simulation, $\mathcal{B} \cup \mathcal{S}$ are generated according to a uniform random distribution. Meanwhile, the possible locations where the RNs can be placed, i.e., $\mathcal{Y}$, are also randomly generated according to a uniform random distribution.

The algorithms are evaluated through python and an extended package for graph theory called networkx [26]. During the simulation, the number of possible locations of RNs $\mathcal{Y}$ is fixed to 200. The algorithms are evaluated by varying the receiver γ_{th}, sensitivity threshold, the number of SNs
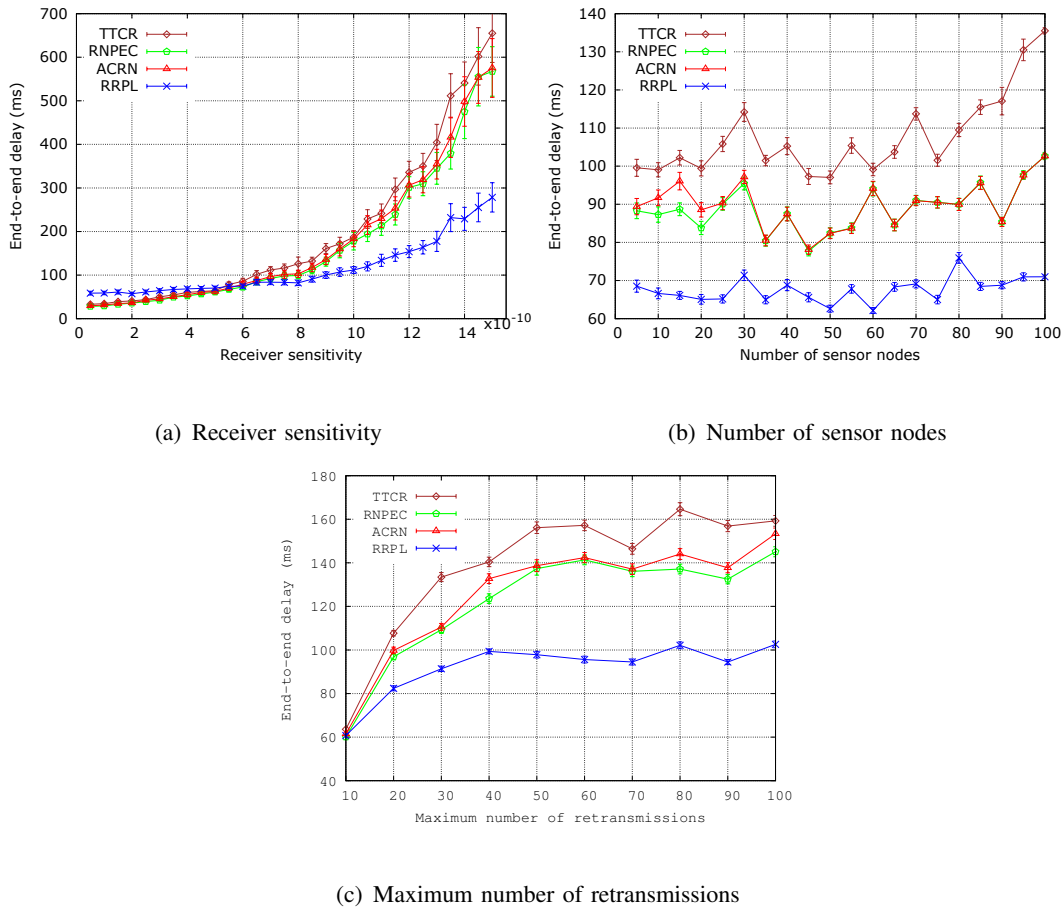
(a) Receiver sensitivity

(b) Number of sensor nodes

(c) Maximum number of retransmissions

Fig. 2. The end-to-end delay measured in terms of the time that a packet spends to travel from $\mathcal{S}$ to $\mathcal{B}$

$\mathcal{S}$ and the maximum number of retransmissions M. We conduct three set of experiments: $i)$ We vary the number of sensor nodes $\mathcal{S}$, and fix γ_{th} to 8×10^{-4} and the number of retransmissions M to 10; $ii)$ We vary γ_{th}, and fix the number of sensor nodes $\mathcal{S}$ and the number of retransmissions M to 10; $iii)$ We vary the number of retransmissions M, and fix γ_{th} to 8×10^{-4} and the number of sensor nodes $\mathcal{S}$ to 50.

Fig. 2 illustrates the performance of the end-to-end delay as a function of γ_{th}, the number of SNs $\mathcal{S}$ and M. The first observation we can make from these figures is that (i) *RRPL* outperforms all other baseline algorithms and (ii) the increase of γ_{th} and M have a negative impact on all the algorithms. From Fig. 2(a), we notice for all algorithms that the end-to-end delay increases with

the receiver sensitivity threshold. In fact, the end-to-end delay have globally the same behavior as the sojourn time. From Section V-C, the sojourn time increases proportionally with the outage probability and the receiver sensitivity threshold. On the other hand, *RRPL* outperforms all the baseline approaches in terms of end-to-end delay. Actually, *RRPL* chooses the links that have the smallest outage probability to sustain the traffic. From Section V-C, the larger the outage probability is, the higher the end-to-end delay becomes. Fig. $2(a)$ shows that *RRPL* does not have a good performance for small values of γ_{th}. This can be explained as follows: *RRPL* uses more RNs $\mathcal{R}$ than *RNPEC*, *ACRN* and *TTCR*, which creates long paths between $\mathcal{S}$ and $\mathcal{B}$. For small values of γ_{th}, all the links succeed to forward the packets without many retransmissions and then the use of long paths by *RRPL* negatively influences the end-to-end delay. From Fig. $2(c)$, we notice that as M increases, the end-to-end delay increases slowly for *RRPL* and quickly for the other algorithms. As discussed in Section V-A, the decrease of the value of M results in a decline of the average number of retransmissions and thus in a drop in the end-to-end delay.

Figures 3 and 4 show the performances of the goodput and network lifetime, respectively, as a function of γ_{th}, the number of SNs $\mathcal{S}$ and M. The results clearly show that *RRPL* outperforms all the baseline approaches in terms of goodput and network lifetime. The use of single-tiered topology by *RNPEC* and *ACRN* affects dramatically the network lifetime in these protocols. Fig. 3 shows that *RRPL* enhances the goodput by more than 50% compared to the baseline approaches. Fig. 4 shows that *RRPL* extends the network lifetime by a factor of two compared to *TTCR* algorithm and by a factor of five compared to *RNPEC* and *ACRN* algorithms. We observe also that γ_{th} has a negative impact on the goodput and the network lifetime, while the increase of M has positive impact on the goodput and negative impact on the network lifetime.

(a) Receiver sensitivity

(b) Number of sensor nodes

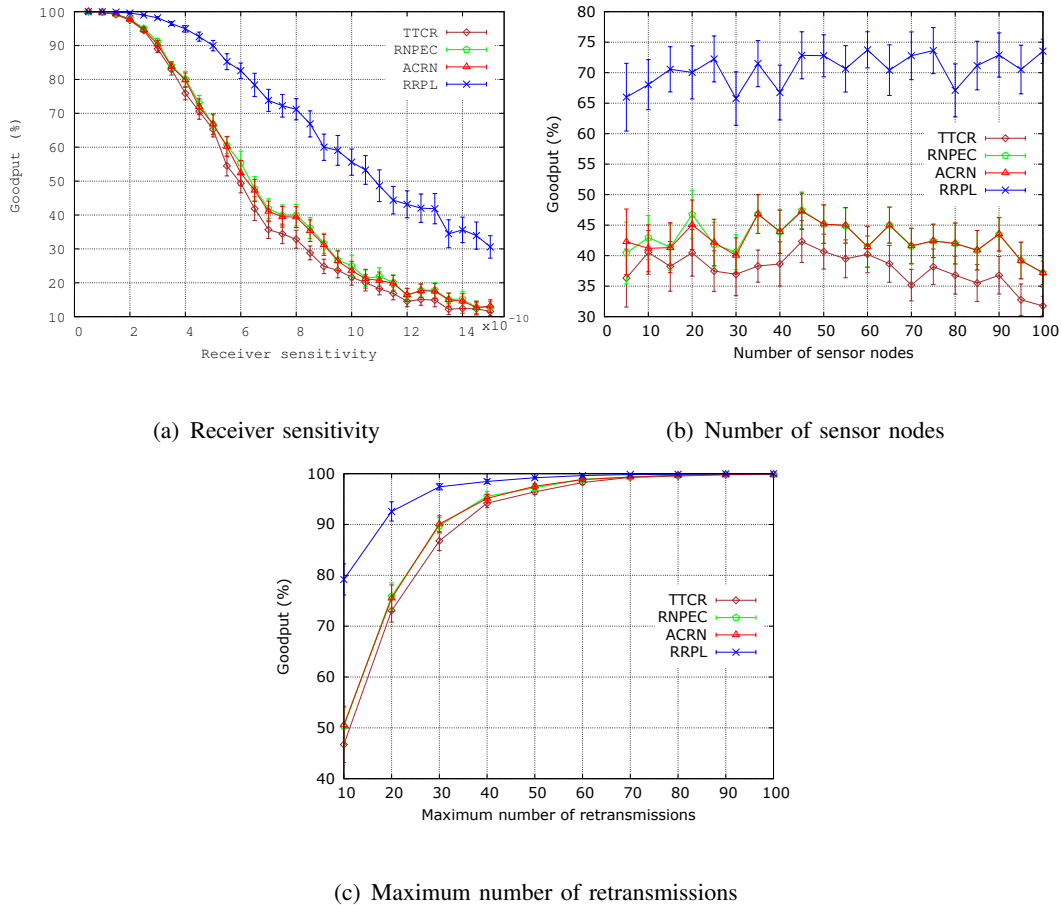(c) Maximum number of retransmissions

Fig. 3. The goodput measured as the number of useful information bits reaching the BSs $\mathcal{B}$

In fact, the increase of the maximum number of retransmissions M leads to the increase of the probability of successful reception and thus the improvement of the goodput. As mentioned in Section V-D, the increase of M results in more energy consumption and consequently leads to a shorter network lifetime. While the increase in the receiver sensitivity threshold γ_{th} leads to a larger outage probability which has a negative impact on the goodput and on the number of retransmissions. The increase in the latter yields more power consumption, which results in a shorter network lifetime.

Fig. 5 shows the cost as a function of γ_{th}, the number of SNs $\mathcal{S}$ and M. The results clearly show that *RNPEC* and *ACRN* outperform *TTCR*, which is in its turn outperforms *RRPL*. This can be explained as follows: In contrast to *RRPL* and *TTCR*, *RNPEC* and *ACRN* form single-tiered

(a) Receiver sensitivity

(b) Number of sensor nodes

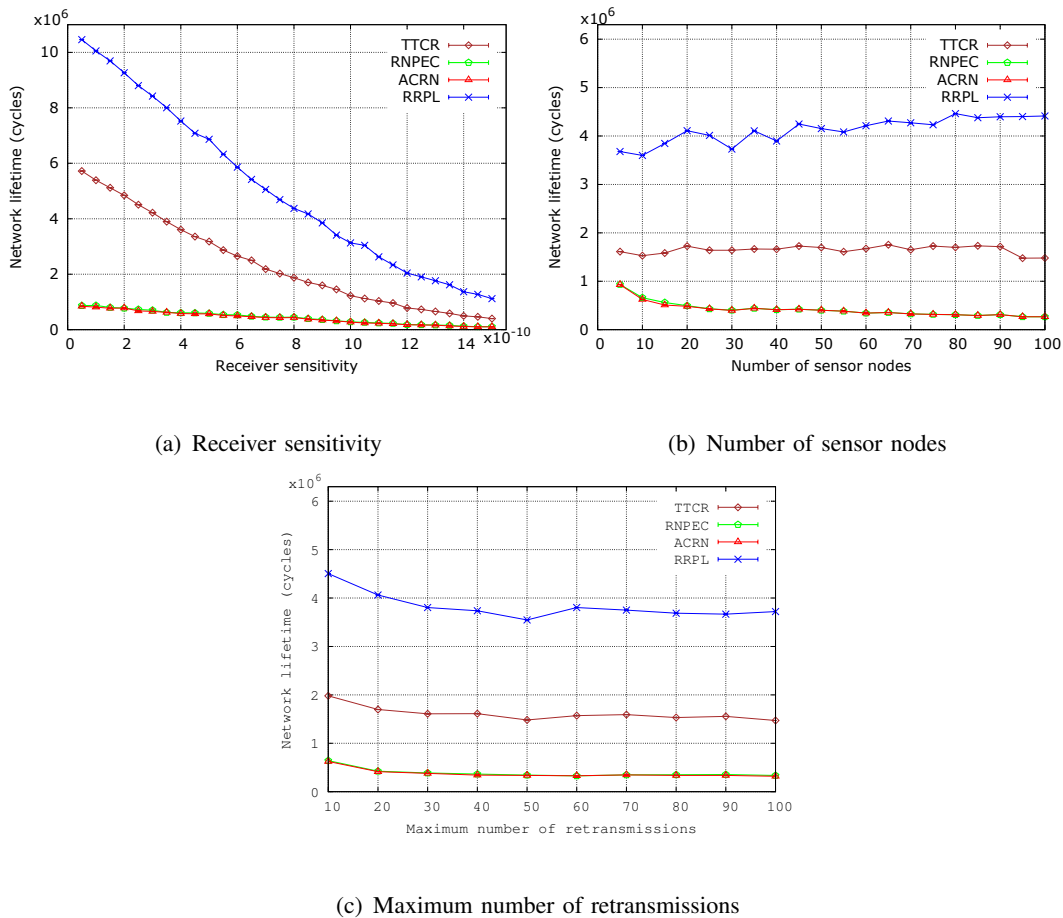(c) Maximum number of retransmissions

Fig. 4. The network lifetime measured as the average time to battery drain out of sensor nodes

topology, where $\mathcal{S}$ participate in the data forwarding, which leads to the reduction of the number

of required relays $\mathcal{R}$. However, the use of single-tiered topology has a negative impact on the

network lifetime as depicted in Fig. 4. While *TTCR* has a lower cost than *RRPL*. In contrast to

RRPL that aims to enhance the QoS while deploying a minimum number of RNs, *TTCR* has

only one objective, which is the reduction of the number of deployed RNs $\mathcal{R}$ while forming

two-tiered topology. From Fig. 5(b), we can observe that the increase on the number of $\mathcal{S}$ has a

positive impact on the cost for *RNPEC* and *ACRN* and a negative impact for *TTCR* and *RRPL*.

For *TTCR* and *RRPL*, the higher the number of SNs $\mathcal{S}$ is, the larger the number of required

RNs for data forwarding becomes. While for *RNPEC* and *ACRN*, the increase of the number of

(a) Receiver sensitivity

(b) Number of sensor nodes

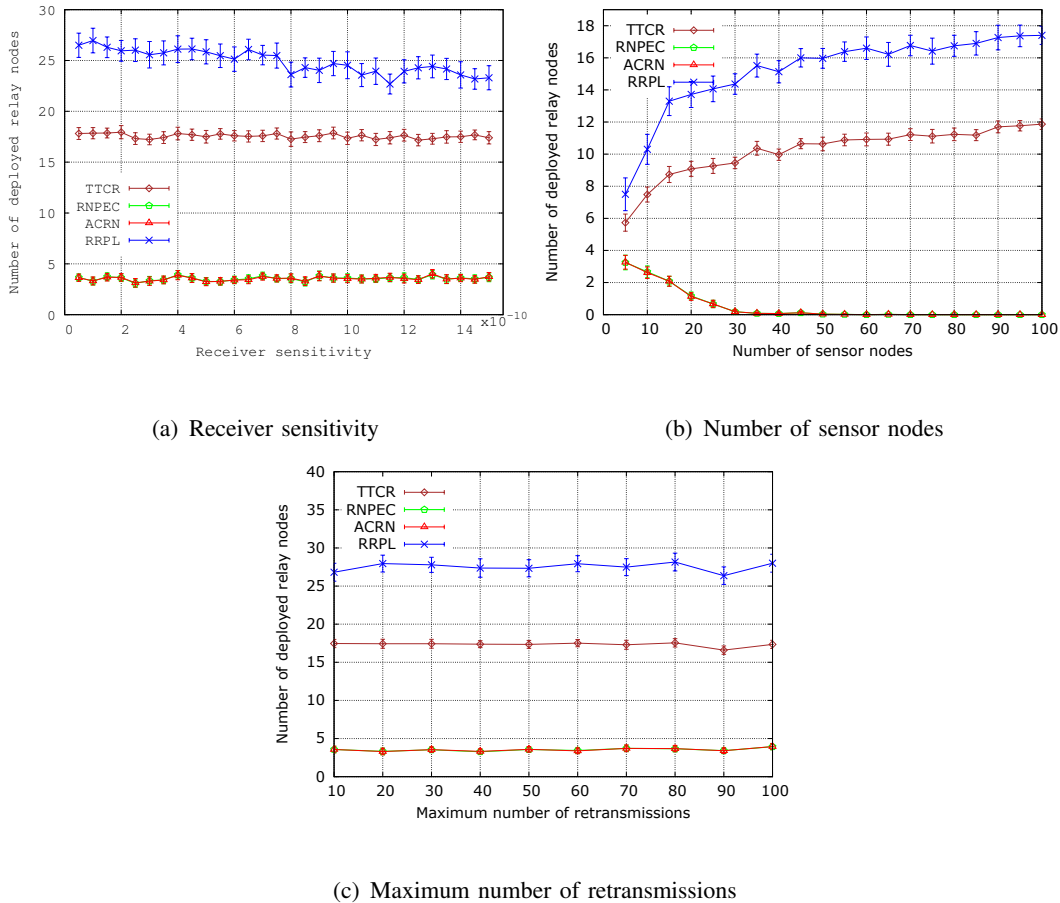
(c) Maximum number of retransmissions

Fig. 5. The cost of solutions measured as the number of relay nodes $\mathcal{R}$ should be added to ensure the connectivity between $\mathcal{S}$ and $\mathcal{B}$.

SNs $\mathcal{S}$ leads to increase the chances of exploiting these added SNs for data forwarding, which reduces the need for extra RNs.

VII. CONCLUSION

To extend the network lifetime, constrained relay node placement has been deemed an effective scheme in the realm of WSNs. In this paper, we have considered the problem of increasing the network goodput and enhancing the QoS while deploying the minimum number of relay nodes in the network. We have devised a solution, named *RRPL*, that exploits the physical model and outage probability to deploy the minimum number of relay nodes that have efficient links for

handling the network traffic. This strategy helped us to increase the goodput and network lifetime and to reduce the data transfer delay. The simulation results have demonstrated the effectiveness of the proposed solution for achieving its design goals. *RRPL* enhanced the baseline approaches with more than 50% in terms of network lifetime, goodpud and data transfer delay. These gains come at a reasonable increase in the number of added relay nodes.

In future work, the effect of shadowing will be taken into account in the physical channel model. The shadowing can be neglected in the case of open field scenarios as assumed in this paper. However, for indoor environment or in industrial workplace, the shadowing effect due to the walls and the equipment is significant and cannot be ignored. Therefore, it is important to take into account the shadowing in such scenarios. The inclusion of the shadowing in the physical model will make our investigation more general. Moreover, enhancements to the *RRPL* algorithm will be proposed to improve the network QoS even in presence of shadowing.

APPENDIX A

PROOF OF THEOREM 1

In this appendix, we derive the proof for the outage probability between two nodes u and v in the network. By definition, the link $u - v$ is in outage if $\mathsf{SINR}_{u,v}$ falls below a threshold level γ_{th} [19]. This event occurs with a probability $\mathcal{P}_{u,v}$. In order to determine an expression for the outage probability $\mathcal{P}_{u,v}$, we need first to compute the CDF of $\mathsf{SINR}_{u,v}$.

First, we recall that the expression of the SINR is given by

$$\mathsf{SINR}_{u,v} = \frac{\gamma_{u,v}}{1 + \sum_{t \in \mathcal{N}}^{t \neq u,v} \gamma_{t,v}}, \tag{A.1}$$

where $\gamma_{u,v}$ and $\gamma_{t,v}$ are exponential random variables with mean values $\bar{\gamma}_{u,v}$ and $\bar{\gamma}_{t,v}$. Using (2)

and (3), the terms $\bar{\gamma}_{u,v}$ and $\bar{\gamma}_{t,v}$ can be computed as

$$\bar{\gamma}_{u,v} = \frac{P_u}{N_0}\left(\frac{d_0}{d_{u,v}}\right)^{\eta} \qquad \text{and} \qquad \bar{\gamma}_{t,v} = \frac{P_t}{N_0}\left(\frac{d_0}{d_{t,v}}\right)^{\eta}. \qquad (A.2)$$

The probability density functions (PDFs) of $\gamma_{u,v}$ and $\gamma_{t,v}$ are given by

$$p_{\gamma_{u,v}}(x) = \begin{cases} \frac{1}{\bar{\gamma}_{u,v}}\exp\left(-\frac{x}{\bar{\gamma}_{u,v}}\right) & \text{if} \quad x \geq 0 \\ \\ 0 & \text{otherwise.} \end{cases} \qquad (A.3)$$

$$p_{\gamma_{t,v}}(x) = \begin{cases} \frac{1}{\bar{\gamma}_{t,v}}\exp\left(-\frac{x}{\bar{\gamma}_{t,v}}\right) & \text{if} \quad x \geq 0 \\ \\ 0 & \text{otherwise.} \end{cases} \qquad (A.4)$$

In order to determine the CDF of the SINR, we need first to compute the PDF of the denominator in (A.1). The term $\sum_{t\in\mathcal{N}}^{t\neq u,v}\gamma_{t,v}$ is a sum of independent non-identically distributed exponential random variables whose PDF can be derived as [27, Eq. (14.5-26)]

$$p_{\Sigma}(x) = \begin{cases} \sum_{t\in\mathcal{N}}^{t\neq u,v}\frac{C_{t,v}}{\bar{\gamma}_{t,v}}\exp\left(-\frac{x}{\bar{\gamma}_{t,v}}\right) & \text{if} \quad x \geq 0 \\ \\ 0 & \text{otherwise.} \end{cases} \qquad (A.5)$$

where $C_{t,v} = \prod_{z\in\mathcal{N}}^{z\neq u,v,t}\frac{\bar{\gamma}_{t,v}}{\bar{\gamma}_{t,v}-\bar{\gamma}_{z,v}}$. Let $Y = 1 + \sum_{t\in\mathcal{N}}^{t\neq u,v}\gamma_{t,v}$. Using (A.5) and the fundamental theorem for the transformation of random variables [28], we can compute the PDF of Y as

$$p_Y(y) = p_{\Sigma}(y-1) = \begin{cases} \sum_{t\in\mathcal{N}}^{t\neq u,v}\frac{C_{t,v}}{\bar{\gamma}_{t,v}}\exp\left(-\frac{y-1}{\bar{\gamma}_{t,v}}\right) & \text{if} \quad y \geq 1 \\ \\ 0 & \text{otherwise.} \end{cases} \qquad (A.6)$$

The random variable $\gamma_{u,v}$ follows an exponential distribution and its CDF is given by

$$F_{\gamma_{u,v}}(x) = 1 - \exp\left(-\frac{x}{\bar{\gamma}_{u,v}}\right). \qquad (A.7)$$

The SINR in (A.1) can be rewritten as $\mathsf{SINR}_{u,v} = \dfrac{\gamma_{u,v}}{Y}$. The outage probability $\mathcal{P}_{u,v}$ can be evaluated as

$$\mathcal{P}_{u,v} = P\left(\mathsf{SINR}_{u,v} \leq \gamma_{\mathrm{th}}\right) = P\left(\frac{\gamma_{u,v}}{Y} \leq \gamma_{\mathrm{th}}\right) = \mathbb{E}_Y\left[P\left(\gamma_{u,v} \leq \gamma_{\mathrm{th}}\ y|Y=y\right)\right] \tag{A.8}$$

$$= \int_1^\infty F_{\gamma_{u,v}}(\gamma_{\mathrm{th}}\ y)p_Y(y)dy = \int_1^\infty \left[1 - \exp\left(-\frac{\gamma_{\mathrm{th}}\ y}{\bar{\gamma}_{u,v}}\right)\right] \sum_{t\in\mathcal{N}}^{t\neq u,v} \frac{C_{t,v}}{\bar{\gamma}_{t,v}} \exp\left(-\frac{y-1}{\bar{\gamma}_{t,v}}\right) dy$$

$$= 1 - \sum_{t\in\mathcal{N}}^{t\neq u,v} C_{t,v} \frac{\bar{\gamma}_{u,v}}{\bar{\gamma}_{u,v} + \bar{\gamma}_{t,v}\gamma_{\mathrm{th}}} \exp\left(-\frac{\gamma_{\mathrm{th}}}{\bar{\gamma}_{u,v}}\right). \tag{A.9}$$

In (A.8), the notation $\mathbb{E}_Y(\cdot)$ stands for the expectation operation with respect to the random variable Y. For any given network topology, we can determine the distance between the different nodes and using (A.2) we can compute the parameters $\bar{\gamma}_{u,v}$ and $\bar{\gamma}_{t,v}$. Finally, utilizing (A.9), the outage probability of any link in the network can be evaluated accurately.

APPENDIX B

AVERAGE NUMBER OF RETRANSMISSIONS

In this appendix, we derive the expression of the average number of retransmissions $\mathbb{E}(T_r)$. First, the expression of $\mathbb{E}(T_r)$ is derived for a value of $M = 2$ and afterwards, we derive $\mathbb{E}(T_r)$ for a general value of M.

For a maximal number of rounds $M = 2$, T_r is a discrete random variable that takes values in $\{1, 2\}$:

- $T_r = 1$, if the transmission of the packet succeeds in the first round. The probability of this event is denoted by $P(S^1)$.

- $T_r = 2$, if the transmission of the data packet fails in the first round, and we have either success or failure in the second transmission round. We denote by $P(F^1, S^2)$ the probability of a reception failure in the first round and a reception success in the second round, while

$P(F^1, F^2)$ stands for the probability of a reception failure in the first and second rounds.

Hence, we can express $\mathbb{E}(T_r)$ for $M = 2$ as

$$\mathbb{E}(T_r) = 1P(S^1) + 2[P(F^1, S^2) + P(F^1, F^2)]. \tag{B.1}$$

Using the following identities

$$P(F^1, S^2) = P(F^1) - P(F^1, F^2) \tag{B.2}$$

$$P(S^1) = 1 - P(F^1), \tag{B.3}$$

we can simplify the expression of the average number of retransmissions in (B.1) as follows

$$\mathbb{E}(T_r) = 1 - P(F^1) + 2[P(F^1) - P(F^1, F^2)] + 2P(F^1, F^2) = 1 + P(F^1). \tag{B.4}$$

For $M \geqslant 2$, we have

$$\mathbb{E}(T_r) = 1P(S^1) + 2P(F^1, S^2) + ... + (M-1) \cdot P(F^1, ..., S^{M-1})$$

$$+ M[P(F^1, ..., F^M) + P(F^1, ..., S^M)]. \tag{B.5}$$

Using the fact that the following equation holds

$$P(F^1, ..., F^{m-1}, S^m) = P(F^1, ..., F^{m-1}) - P(F^1, ..., F^m), \tag{B.6}$$

for $m = 2, ..., M$, the average number of retransmissions can be written as

$$\mathbb{E}(T_r) = 1P(S^1) + 2P(F^1, S^2) + ... + (M-1)P(F^1, ..., S^{M-1}) + MP(F^1, ..., F^{M-1})$$

$$= 1 - P(F^1) + 2[P(F^1) - P(F^1, F^2)] + (M-1) \cdot [P(F^1, ..., F^{M-2}) - P(F^1, ..., F^{M-1})]$$

$$+ MP(F^1, ..., F^{M-1}) = 1 + \sum_{m=1}^{M-1} P(F^1, ..., F^m). \tag{B.7}$$

REFERENCES

[1] D. Yang, S. Misra, X. Fang, G. Xue, and J. Zhang, "Two-tiered constrained relay node placement in wireless sensor networks: computational complexity and efficient approximations," *IEEE Transactions on Mobile Computing*, vol. 11, no. 8, pp. 1399–1411, aug 2012.

[2] S. Misra, N. E. Majd, and H. Huang, "Approximation algorithms for constrained relay node placement in energy harvesting wireless sensor networks," *IEEE Trans. on Computers*, vol. 63, no. 12, pp. 2933–2947, Dec. 2014.

[3] D. Djenouri and M. Bagaa, "Energy harvesting aware relay node addition for power-efficient coverage in wireless sensor networks," in *IEEE ICC*, 2015, pp. 86–91.

[4] J. Bredin, E. Demaine, M. Hajiaghayi, and D. Rus, "Deploying sensor networks with guaranteed fault tolerance," *IEEE/ACM Transactions on Networking*, vol. 18, no. 1, pp. 216–228, feb 2010.

[5] X. Cheng, D.-Z. Du, L. Wang, and B. Xu, "Relay sensor placement in wireless sensor networks," *Wireless Networks*, vol. 14, no. 3, pp. 347–355, jan 2007.

[6] G.-H. Lin and G. Xue, "Steiner tree problem with minimum number of Steiner points and bounded edge-length," *Information Processing Letters*, vol. 69, no. 2, pp. 53–57, jan 1999.

[7] S. Misra, S. D. Hong, G. Xue, and J. Tang, "Constrained relay node placement in wireless sensor networks: formulation and approximations," *IEEE/ACM Trans. on Networking*, vol. 18, no. 2, pp. 434–447, Apr. 2010.

[8] H. A. Hashim, B. Ayinde, and M. Abido, "Optimal placement of relay nodes in wireless sensor network using artificial bee colony algorithm," *Journal of Network and Computer Applications*, vol. 64, pp. 239 – 248, 2016.

[9] R. Liu, I. J. Wassell, and S. K., "Relay node placement for wireless sensor networks deployed in tunnels," in *IEEE 6th International Conference on Wireless and Mobile Computing, Networking and Communications, WiMob 2010*, 2010.

[10] D. Wu, D. Chatzigeorgiou, K. Youcef-Toumi, S. Mekid, and R. Ben-Mansour, "Channel-aware relay node placement in wireless sensor networks for pipeline inspection," *IEEE Trans. on Wireless Commun.*, vol. 13, no. 7, pp. 3510–3523, Jul. 2014.

[11] Q. Chen, Z. Hu, Y. Chen, V. Grout, D. Zhang, H. Wang, and H. Xing, "Improved relay node placement algorithm for wireless sensor networks application in wind farm," in *IEEE International Conference on Smart Energy Grid Engineering (SEGE)*, Aug. 2013.

[12] C. Ma, W. Liang, and M. Zheng, "Set-covering-based algorithm for delay constrained relay node placement in wireless sensor networks," in *IEEE ICC 2016*, 2016.

[13] M. Nikolov and J. Zygmunt Haas, "Relay placement in wireless networks: minimizing communication cost," *IEEE Trans. on Wireless Commun.*, vol. 15, no. 5, pp. 3587–3602, May 2016.

[14] Z. Zheng, L. X. Cai, R. Zhang, and X. S. Shen, "RNP-SA: Joint relay placement and sub-carrier allocation in wireless communication networks with sustainable energy," *IEEE Trans. on Wireless Commun.*, vol. 11, no. 10, pp. 3818–3828, Oct. 2012.

[15] A. Efrat, S. P. Fekete, P. Gaddehosur, J. Mitchell, V. Polishchuk, and J. Suomela, "Improved approximation algorithms for relay placement," in *ESA (Ann. European Symp. Algorithms)*. Springer Berlin Heidelberg, 2008, pp. 356–367.

[16] W. Zhang, G. Xue, and S. Misra, "Fault-tolerant relay node placement in wireless sensor networks: problems and algorithms," in *IEEE INFOCOM*, 2007, pp. 1649–1657.

[17] E. Lloyd and G. Xue, "Relay node placement in wireless sensor networks," *IEEE Transactions on Computers*, vol. 56, no. 1, pp. 134–138, jan 2007.

[18] A. Chelli, M. Bagaa, D. Djenouri, I. Balasingham, and T. Taleb, "One-Step Approach for Two-Tiered Constrained Relay Node Placement in Wireless Sensor Networks," *IEEE Wireless Commun. Letters*, vol. 5, no. 4, pp. 448–451, 2016.

[19] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels*, 2nd ed. New Jersey, USA: John Wiley & Sons, 2005.

[20] A. F. Molisch, *Wireless Communications*. Wiley-IEEE Press, 2005.

[21] L. Kou, G. Markowsky, and L. Berman, "A fast algorithm for Steiner trees," *Acta Informatica*, vol. 15, no. 2, pp. 141–145, 1981.

[22] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products, Seventh Edition*. Academic Press, 2007.

[23] W. Chan, T.-C. Lu, and R.-J. Chen, "Pollaczek-Khinchin formula for the M/G/1 queue in discrete time with vacations," *IEE Proceedings - Computers and Digital Techniques*, vol. 144, no. 4, pp. 222–226, Jul. 1997.

[24] A. Chelli, E. Zedini, M.-S. Alouini, J. R. Barry, and M. Pätzold, "Performance and delay analysis of hybrid ARQ with incremental redundancy over double Rayleigh fading channels," *IEEE Trans. on Wireless Commun.*, vol. 13, no. 11, pp. 6245–6258, Nov. 2014.

[25] G. Y. Li, Z. Xu, C. Xiong, C. Yang, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient wireless communications: tutorial, survey, and open issues," *IEEE Trans. on Wireless Commun.*, vol. 18, no. 6, pp. 28–35, Jul. 2011.

[26] Networkx, "http://networkx.lanl.gov."

[27] J. G. Proakis, *Digital Communications*. McGraw-Hill, 2008.

[28] A. Papoulis and S. Pillai, *Probability, random variables, and stochastic processes*, 4th ed. McGraw-Hill, 2002.