A flow-based optimization model for throughput-oriented relay node placement in Wireless Sensor Networks

Eduardo Feo Flushing
Dalle Molle Institute for
Artificial Intelligence (IDSIA)
Lugano, Switzerland
eduardo@idsia.ch

Gianni A. Di Caro
Dalle Molle Institute for
Artificial Intelligence (IDSIA)
Lugano, Switzerland
gianni@idsia.ch

ABSTRACT

This work addresses the relay node placement problem in wireless sensor networks. We consider a scenario in which all sensor nodes stream data towards sink nodes. Additional relay nodes can be placed with the aim of optimizing overall network performance in terms of data throughput. We formalize the problem with a linear, mixed integer mathematical programming model. We include a number of constraints and penalties to closely model the wireless environment. When relay nodes can be placed anywhere, we define their possible locations using a discrete grid. Model solutions specify both where to place the relays and the paths for routing. Through extensive simulation experiments, we compare model solutions against a state-of-the-art dynamic routing protocol to assess the quality of the routes, and against a relay node placement heuristic to evaluate relay positioning. To tackle the computational complexity, we also propose and study the effect of different strategies for determining the grid resolution.

Additionally, an experimental validation carried out in a real testbed shows that the computed solutions clearly increase network performance by enabling the reception of larger number of data packets at the sinks and determining a fair QoS distribution among the nodes. Finally, we propose an on-line application, in which the model is built and solved on-demand, to adapt to changes in traffic patterns induced by external events, and the results are rapidly spread throughout the network and used to modify relay node positioning and routing paths.

Categories and Subject Descriptors

C.2.1 [**Network Architecture and Design**]: Wireless Communications; C.2.3 [**Network Operations**]: Network management

Keywords

Relay Node Placement, Wireless Sensor Networks, Integer Programming, Optimization, Multi-hop, Routing

SAC'13 March 18-22, 2013, Coimbra, Portugal.
Copyright 2013 ACM 978-1-4503-1656-9/13/03 ...$15.00.

1. INTRODUCTION

A *wireless sensor network* (WSN) consists of a set nodes that are equipped with sensing and limited processing capabilities, and can locally communicate with each other through a wireless medium [1]. The *sensor nodes* (SNs) composing a WSN are usually inexpensive and low-powered, such as they can be deployed in large numbers to provide monitoring and sensing services for long time periods. In typical applications, the data generated by the sensor nodes need to be transmitted to and aggregated and processed at *base stations* (BSs). The general model for the forwarding of the data from SNs to BSs is based on the definition and use of *multi-hop routing paths*.

Since a WSN can operate for relatively long times and/or it can be embedded in dynamic or hostile environments, a core issue in WSNs is the definition of effective strategies for the time maintenance of network operativity and/or for its adaptation to external or internal changes. In this direction, a wealth of research has considered the use of special nodes, referred to as *relay nodes* (RNs), that can be deployed and added to the WSN after the network has been put in place. RNs can be positioned at precise locations by hand, or they can be part of a mobile robotic unit, such that they can be deployed autonomously or on-demand.

Possible uses of RNs include the provisioning of connectivity and the enhancement of the network response (e.g., in terms of throughput, sensing, and lifetime), where and when needed. For instance, RNs can be used to provide connectivity when the network becomes partially disconnected from the BSs because of factors such as energy depletion, node failures, interaction with a hostile environment, etc. In other scenarios, such as the tracking of moving targets and the monitoring of dynamic environments, RNs can help the network to locally adapt to the external changes.

In this paper we address the *relay node placement for performance enhancement* problem (RNP-PE in short), defined as follows. Given a set of locations where it is feasible to deploy a restricted number of available RNs, the objective is to select from this set the locations where the RNs can be positioned in order to improve throughput and end-to-end packet delays for the data gathered at BSs. When the network is partially connected (i.e., some SNs are not be able to send data to BSs), the deployment of RNs is aimed both at bringing connectivity and optimizing global network performance. Although our primary objective is determining the physical locations at which RNs should be placed, the solution also specifies the way these RNs should be used. This specification comes in the form of *optimal routing paths* from

SNs to BSs to forward data flows. We present the RNP-PE as a flow-based *linear, mixed integer mathematical program* (MIP). The formulation includes a number of constraints and penalty components, aimed at closely modeling the specific characteristics of the wireless environment, as well as a number of heuristics, aimed at speeding up the computations. The model is solved to optimality using a standard solver, finding the best locations where to position the available RNs and the paths to route data flows.

In real-world applications, the positioning of RNs becomes influential when the network environment, that is, the network topology and traffic loads, are dynamically changing. For instance, traffic congestion can be caused by an increase in the number of data packets generated by the sensor nodes close to a recent event (e.g., as in target detection in surveillance applications, or in the occurrence of natural events in environment monitoring), or because the failures of some SNs have disrupted some of the routes in use. To this end, the RNP-PE model can be used to reactively reposition groups of (possibly mobile) RNs in a real-time fashion, timely adapting to the changes in the network environment.

The main scientific contribution of this work, compared to previous MIP models for relay node placement, is to present an approach that addresses both network performance enhancement and connectivity issues, and explicitly includes in the model most of the critical aspects of interference and congestion in wireless environments. Moreover, we validate the efficacy of the proposed model through both simulations and experiments using real wireless sensor devices. We also present an on-line application of the model which allows to adapt the network to external changes and improve the quality of service in real-time.

2. RELATED WORK

In recent years, a number of studies in WSNs have considered the *relay node placement problem* (RNP) under different requirements and objectives. Most of the existing work has focused on the deployment of RNs for the provisioning of: *connectivity* [19, 3, 4, 20, 6, 18, 21, 13, 24], *extended network lifetime* [14, 25], *energy-efficient or balanced data gathering* [8, 22, 7], and *survivability and fault tolerance* [26, 27, 15, 21, 24, 12].

Studies based on connectivity requirements aim to place a minimum number of RNs in order to achieve different goals. In the *Connected* RP problem, the objective is to make the induced network topology globally connected, assuming that the nodes were initially partially disconnected. In the *Survivable* or *Fault-tolerant* approach, the goal is to ensure that the network remains connected in the presence of $k \geq 1$ failures. For the Connected RNP, the problem can be described as deploying a minimum number of RNs such that, for each pair of SNs, it exists a path consisting of RNs and/or SNs and each hop of the path is no longer than the common transmission range of the SNs and the RNs. Lin and Xue [19] formulated the problem as the *Minimum Steiner Tree* with minimum number of Steiner points and bounded edge length. They proved that the problem is NP-hard and proposed a minimum spanning tree 5-approximation algorithm. Following this work, several other authors tackled the Connected and the Survivable RNPs by proposing different approximation algorithms. Although our problem shares some similarities with these works with respect to the provisioning of connectivity when needed, we do not focus on optimality in terms of number of RNs. We aim to deploy RNs to provide connectivity, when needed, and at the same time to enhance the overall network performance. The definition of the optimal number of RNs to use is a by-product of our approach, which selects the number of effectively useful RNs out of a defined maximum number of RNs potentially made available. Also, none of the previously mentioned works considered the RNP from a real-world point of view, that is, taking into consideration physical network properties and/or wireless channel characteristics.

The approaches focusing on performance metrics other than connectivity and survivability are reviewed in the following, where for each approach we single out some core differences, in terms of objectives and modeling with the work which we present in this paper. Falck *et al.* [8] have considered the RNP in the context of balanced data gathering. They presented the problem of finding an optimal routing as a linear program, but with the objective of achieving load balancing. The placement of relay nodes is approximated by adding relays one at a time. To the contrary, we directly include the placement in our mathematical model, and jointly solve both problems (i.e., optimal routing and placement). Patel *et al.* [22] examined the joint problem of deploying SNs, RNs, and BSs on a set of feasible locations and finding bandwidth-constrained energy-efficient routes with guaranteed coverage, connectivity, bandwidth, and robustness. They also make the use of a linear program formulation. These authors have considered the objective of network performance in the form of maximizing network utilization when RNs can be deployed only in a set of feasible sites. Although we also consider a numerable set of sites where RNs can be positioned, we propose strategies to limit its size and tackle the computational complexity of the problem. Kashyap *et al.* [16] studied the placement of RNs with the goal of reducing maximum link load for a given traffic imposed on the sensors. They also considered a flat architecture, called backbone network, but without base stations. Instead, they considered demands between any pair of nodes in the network (called profile entries). Their model restricted the placement of the relays to the lines joining the backbone nodes, determining the minimum number of relays needed to link two nodes.

Ergen and Varaiya [7] considered the problem of determining optimal locations for RNs together with optimal energy provisioning, such that the network operates for the desired lifetime with minimum energy expenditure. These authors considered a non-linear programming model and established a set of possible locations for the RNs based on a grid partitioning. The work mostly focuses on energy-efficiency, such that the quality of the solutions in terms of network performance is not really investigated, and considers a simplistic radio propagation model. In [23], the authors studied the joint problem of placing relay nodes and scheduling node transmissions in the presence of controlled mobility. The approach aims to maximize the lowest weighted throughput among the nodes in the network. Only star topologies are considered, in which nodes communicate to each other only through a relay node. In our work, we aim to maximize the overall network throughput by the creation of data routes of higher quality and the reduction of traffic congestion through the positioning of relay nodes, and we consider multi-hop topologies. Capone *et al.* [2] proposed a network flow based model for the optimal routing in wireless mesh networks addressing TDMA networks and focusing on traffic scheduling. Due to their simplicity, CSMA based MAC protocols are preferred in WSNs. Therefore, our model explicitly considers this type of network. Finally, Wang *et al.* [25] studied the deployment of RNs to maximize network lifetime in two-tiered WSNs with a single base station. We consider general flat topologies and the presence of multiple BSs, to address large-scale scenarios.

3. SYSTEM MODEL

We model the WSN as a set of SNs and BSs located in a set of known positions $\mathcal{S}$ and $\mathcal{B}$, respectively. SNs both generate and forward data packets towards one of the BSs in multi-hop fashion (a data flow can be split over multiple paths). We assume that the expected data generation characteristics of each SN are known. All nodes communicate with each other within the *communication range* $\mathbf{r}$.

A set of K RNs is also available, their role is to forward data received from other nodes. The placement of node relays is restricted to a *numerable set of candidate locations* $\mathcal{R}$. This set may be pre-defined by the application scenario (as seen later in the experimental evaluation), or determined using a a *2D grid* covering the area.

We formalize the problem of RN deployment by a *linear, mixed integer mathematical programming* (MIP) model based on a *minimum cost flow* formulation that includes a number of additional constraints and penalty components. The model is incrementally constructed in the next section.

3.1 MIP Formulation

Let $G = (V, E)$ be a connected digraph representing a WSN, where $V = \mathcal{N} = \mathcal{S} \cup \mathcal{B} \cup \mathcal{R}$ is the set of all node positions , and E is the set of communication links. $\gamma : E \mapsto \Re$ is a *link cost* function, and $\tau : \mathcal{S} \mapsto \Re$ is a data generation (*traffic load*) function, expressed as the data per second generated by an SN. In the following, we measure τ in terms of *flow units*, f_{unit}, expressed as bytes/sec.

Data flows and relay positions define the two sets of *decision variables*. The *flow variables* f_{ij} denote the amount of flow through link (i, j), that is, the traffic load from node n_i to node n_j, located at positions $i,\ j \in \mathcal{N}$ respectively. f_{ij} values are expressed in *flow units*. The *binary positional variable* y_i indicates whether location $i \in \mathcal{R}$ (a point in the 2D grid) is being used to circulate flow or not. When y_i is set to 1 in a solution, an RN is to be positioned at the corresponding grid location. A full solution specifies both flows and relay positions. The SN-to-BS routes are defined in the *routing-tree* induced by the set $\{(i,\ j) \in E \mid f_{ij} > 0\}$. The formulation of the minimum cost flow problem with RN placement is:

$$\min \quad \text{RNP-PE} = \sum_{(i,\ j) \in E} \gamma_{ij} f_{ij} \tag{1}$$

$$\text{subject to:} \quad \sum_{(i,\ j) \in E} f_{ij} - \sum_{(j,\ i) \in E} f_{ji} = \begin{cases} \tau_i & \text{if } i \in \mathcal{S}, \\ 0 & \text{if } i \in \mathcal{R} \end{cases} \tag{2}$$

$$\sum_{i \in \mathcal{B}} \sum_{(j,\ i) \in E} f_{ji} = \sum_{k \in S} \tau_k \tag{3}$$

$$y_i = 1 \iff \sum_{j \in \mathcal{N}} f_{ji} > 0 \quad \forall i \in \mathcal{R} \tag{4}$$

The number of available RNs can be limited to K by adding the following constraint: $\sum_{i \in \mathcal{R}} y_i \leq K$.

Given that the optimal solution can correspond to a number of RNs $k < K$, we include a penalty factor in the objective (1) to favor the use of a minimal amount of RNs. Any solution using n relays needs to provide a minimal gain $\hat{R}$ compared to the same solution obtained using $n-1$ relays:

$$\min \ \text{RNP-PE} = \sum_{(i,\ j) \in E} \gamma_{ij} f_{ij} + \hat{R} \sum_{i \in \mathcal{R}} y_i, \tag{5}$$

where $\hat{R}$ is a parameter that can be adjusted according to the problem instance (e.g., relay node availability, economic cost). In the tests of Section 4, $\hat{R}$ is set to one flow unit.

Shared wireless channels in WSNs are necessarily *bandwidth-limited*. To specify this as a constraint, we use the notion of *link capacity* L_{cap}, which is the nominal amount of data (bytes/sec) that can be transmitted by a wireless link in the network (assuming the same capacity for all links). We express it in flow units. Using link capacity, the following constraint formulates bandwidth limitations:

$$\sum_{(i,\ j) \in E} f_{ij} + \sum_{(j,\ i) \in E} f_{ji} \ \leq \ L_{cap}, \quad \forall i \in \mathcal{N} \tag{6}$$

For a node n, the *routing in-degree* is the number of n's neighbors using n to relay data. Because of shared medium and contention access, this number strongly impacts on the effective node capacity and network load balancing. For this reason, the following constraint adds to the model a restriction on the maximum routing in-degree allowed (b is a binary auxiliary variable, $D = 6$ in the experiments):

$$b_{ij} = 1 \iff f_{ij} > 0 \ \forall i,\ j \in \mathcal{N}, \quad \sum_{(j,\ i) \in E} b_{ji} \leq D \ \forall i \in \mathcal{S} \tag{7}$$

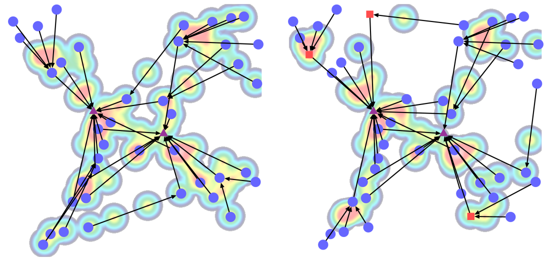
To minimize wireless interference and produce balanced routing trees, which allow balanced energy depletion, we need to setup *minimally interfering flow paths.* We enforce this by introducing in the objective function a penalty component based on the *maximum local flow*, $\bar{F}_{max}$, defined as the maximum amount of flow that can circulate within a disk of radius r centered in an SN. The calculation of the flow circulating within the r-disk of an SN i, requires to sum up the outgoing flows from all i's neighbors:

$$p_i = 1 \iff \left(\sum_{(i,\ j) \in E} \sum_{(j,\ k) \in E} f_{jk} \geq \bar{F}_{max} \right) \tag{8}$$

where p_i is a binary variable that takes value 1 when the flow through the i's r-disk violates the maximum allowed amount.

In order to use p to include in the objective function the penalty, and give it a reasonable weight, we first derive an estimate of the order of magnitude of the objective function value without any penalties, Eq. (1), as follows. Given a link cost function γ, we can determine an upper bound on the *maximum link cost*, γ_{max}. A lower bound on the path length from an SN i to a BS can be defined by considering the minimum Euclidean distance between i and BS divided by the transmission range. Therefore, the following expression is an estimate of the optimal solution value of RNP-PE:

$$\hat{F} = \left(\sum_{i \in \mathcal{S}} \min_{j \in \mathcal{B}} \left(\frac{|i,\ j|}{r} \right) \tau_i \gamma_{max} \right). \tag{9}$$



(a) (b)

Figure 1: Effect of congestion and its mitigation through the use of relay nodes.

Using $\hat{F}$ and p, the penalty factor for the violation in maximum local flow is therefore included in the objective function (1), that, together with all the other penalties become:

$$\min\ \text{RNP-PE} = \sum_{(i,\ j)\in E} \gamma_{ij} f_{ij} + \hat{R}\sum_{i\in\mathcal{R}} y_i + \alpha\sum_{i\in\mathcal{S}} p_i\hat{F}. \quad (10)$$

The parameter α weighs the penalty in the objective function. In the experiments we set $\alpha = 0.1$. Figure 1 shows an example of the solutions obtained with and without this element in the model. Red spots in the heat map correspond to links and/or nodes which suffer of high rate of packet losses. In this example, we can appreciate the use of RNs to deviate data traffic and leverage the congestion level of the network. In the example, this reduction of congestion resulted in an increase of 5% in network delivery ratio.

3.1.1 *Definition of link costs*

So far, *link costs* were generically defined as a mapping γ. It is common to define link costs in terms of hop count, in which all links have the same, unit cost. In our case, adopting such a way of proceedings, would be equivalent to minimize the number of links used by each flow, that is, to find the shortest routing trees. However, it is well understood that in WSNs the selection of links following the min-hop rule ignores many critical aspects of the network environment that strongly impact on the overall performance [11]. Therefore, in order to achieve a good trade-off between hop-count and link reliability, we adopted as link cost function $\gamma_{ij} = 1 + w_c c_{ij}$. Where $c : E \mapsto [0, 1]$ is a *link quality estimator* ($c_{ij} < c_{kl}$ means that link (i, j) has a better quality than (k, l)). In the experiments of Section 4, to derive the probability of a successful transmission we used a link quality estimator c combining an interference and a MAC model [9]. However, any other sound estimator could be used.

The weight w_c defines the relative importance between hop-count and link quality estimation. Defining a good trade-off between these two aspects is in general a difficult task. In 4 we report experiments for $w_c = 1.0$, a value which was identified as reasonably good through empirical testing.

3.2 Candidate locations for relay nodes

In certain applications, RNs can be placed anywhere in the field. However, to reduce the complexity of the relay placement problem, we restrict the set of feasible locations to a discrete set, corresponding to grid locations over the network area. Grid points are placed at a regular distance Δ from each other (see Figures 2a and 2b). A small value of Δ defines a large number of candidate RN positions and may lead to better quality solutions, at the expenses of increased computational complexity.

We propose two heuristics to optimize the trade-off between number and quality of candidate locations, and computational complexity. The first heuristic consists in calculating the *convex-hull* of the locations of SNs and BSs and considering as candidate locations for RN placement only those falling within the convex-hull. The rationale of the heuristic is related to the observation that, in general, the optimal placement of the RNs is expected to be found in locations close to the the straight line connecting SNs. It is therefore reasonable to prune from the set of candidate locations those that are outside the convex-hull. Figure 2c shows an example of application of the heuristic.

The second heuristic is based on a *locally adaptive* strategy. It uses different values of Δ in different regions. Larger values of Δ are used in regions with high or low node density, while small values of Δ are used in regions with intermediate node density. The rationale behind this strategy is that in high density regions the degree of connectivity is expected to be also high, therefore RNs may not be needed. In regions with low node density, the selection of RNs will likely serve to connect nodes from adjacent sub-regions, therefore a high precision in positioning is not expected to be critical. For intermediate node densities, no reasonable assumptions can be made and a small value of Δ should be used. In order to implement this strategy, the sensing field is divided uniformly in D sub-regions. Two parameters, Δ_U and Δ_L specify the upper and lower bounds for Δ. δ_i $(i = 1 \dots D)$ and δ_T are, respectively, the node density of each sub-region and of all field. The value $\frac{\delta_T}{D}$, which would be the expected value of δ_i if nodes were deployed uniformly, is denoted as δ_r. The range $[\Delta_L, \Delta_U]$ is sampled in k values $\{\Delta_L = \Delta_1 < \Delta_2 < \dots < \Delta_{k-1} < \Delta_k = \Delta_U\}$, which represent the possible values for Δ_i. The value of Δ_i for each sub-region i is equal to $\Delta_{\widehat{k}}$, where $\widehat{k}$ is defined by:

$$\widehat{k} = \min\left(k, \lceil k \times \left(\frac{|\delta_r - \delta_i|}{\delta_r}\right)\rceil\right).$$

Low densities, $\delta_i \approx 0$, and high densities, $\delta_i \gg \delta_r$, will result in $\widehat{k} = k$, that is, in larger values for $\Delta_i = \Delta_U$. Densities close to δ_r correspond to values $\Delta_i \approx \Delta_L$.

The combined application of the two heuristics is illustrated in Figure 2d.

4. SIMULATION RESULTS

With the purpose of covering a wide set of different application scenarios, we generated 1500 different networks embedded in areas ranging from 50×50 m^2 to 75×75 m^2. The number of SNs is varied from 50 up to 225, and the number of BSs from 12 to 56. The communication range is set to 10 m. Networks are generated with different topological characteristics: uniform, clustered, and small world. Traffic load is modeled using, for each SN, a periodic packet generation with a data rate ranging from 320 to 960 bytes/sec. For each instance 30 runs were executed to account for randomness. In the plots we report the *empirical cumulative distribution function* (indicated with CDF) for *packet delivery ratio* considering the results over all 1500 instances.

To solve the MIP model we used the CPLEX® solver (default settings) and 1 hour as maximum solving time. The ob-

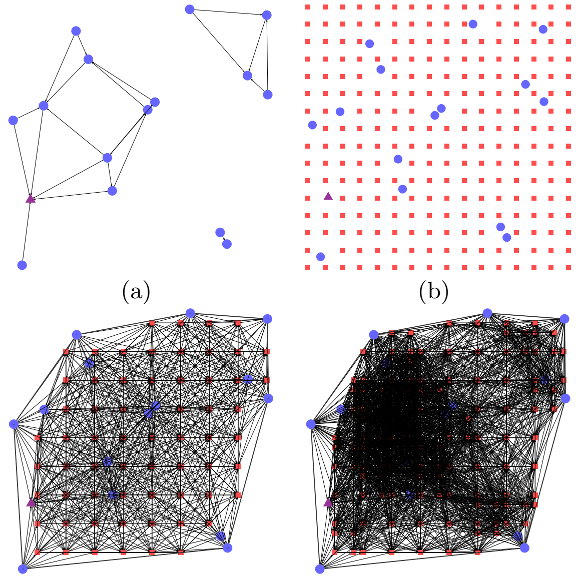


Figure 2: Candidate locations for relays: (a) Network with $|\mathcal{S}| = 15$ sensor nodes and $|\mathcal{B}| = 1$ base station; (b) Squared grid for relay positioning, $\Delta = 1$m, $|\mathcal{R}| = 256$; (c) Grid from convex-hull pruning, $\Delta = 2$m, $|\mathcal{R}| = 151$, $|E| = 3685$; (d) Grid from adaptive strategy and convex-hull, $\Delta = [2\text{m-}5\text{m}]$, $|\mathcal{R}| = 139$, $|E| = 3435$.

tained solutions were tested in simulation using TOSSIM [17].

4.1 Relay node placement for performance

To evaluate the efficacy of the model on deploying RNs with the purpose of enhancing network performance, we have selected only *connected* topologies in the generated set. The values for K are specified in % of the total nodes (SNs and BSs) in the initial topology. Selected values for K are 5%, 10% and 15%. For each value of K, we solve the problem instances and evaluate the provided solution through simulation. Figure 3a shows the network delivery ratio CDF considering different values of K. The value $K = 0$ means that the instances were solved without any relay, and serve as a comparison to assess the performance improvement obtained by deploying RNs. The performance gain obtained by RNs deployment is quite evident once compared to the case $K = 0$. On the other hand, it is not always clear whether adding too many relays can always improve performance.

4.2 Quality of routing paths

The solution of the MIP model provides the number of relay nodes to be used with their positions, as well as the paths to route data flows (flow variables f_{ij}). To evaluate the quality of the obtained routing paths, we compare the network performance (network delivery ratio) obtained using these paths with the performance obtained using the *Collection Tree Protocol* (CTP) [10]. The comparison is not aimed at showing which one between the two approaches is better, but rather at validating the fact that the paths computed through the model are indeed 'good'.

CTP is a state-of-the-art routing and data collection protocol distributed with TinyOS/TOSSIM. By default, the forwarding engine of CTP requests packets acknowledgments and performs retransmissions when they are not received. As we considered only best-effort transmissions and we did not include the use of packet retransmissions in model evaluation, we disabled this functionality in CTP in order to build a fair comparison.

Figure 3b shows that our model clearly outperforms CTP in terms of delivery ratio (and also for end-to-end delay), in spite of the fact that our paths are statically assigned, while CTP define them in an adaptive way.

4.3 Relay node placement for connectivity

Apart from the evaluation of our model in the context of network performance, as done in the previous section, we want to assess the quality of our approach in determining good locations for RNs also in situations in which their deployment is mainly guided by the need of connectivity. At this aim, we consider *partially connected* instances and use the *Minimum Spanning Tree Heuristic* (MST) of Lin [19] to determine the RN positions to achieve connectivity. The performance of our model compared to the MST heuristic is shown in Figure 3c based on simulation experiments. Our model clearly outperforms the MST (note that, based on the RN positions output by the heuristic, we define the routing paths using our model). The main reason is likely the lack of notion of network paths in the MST heuristic, in which the placement is focused on minimizing the number of RNs without explicitly taking into account the quality of the routes from SNs to BSs. It is important to point out that the comparison with the MST is not really aimed at showing that our model can define an RN placement that can result in a better performance (which is expected), but rather in validating the goodness of the computed locations also with respect to the connectivity objective.

4.4 Effect of grid heuristics and resolution

Figure 4a illustrates the effect of different grid resolutions and of the combination of adaptive and convex-hull heuristics. The figure shows the percentage of cases in which the use of the indicated resolution resulted in the best performance during the simulation experiments. As expected, increasing the density of RN locations (i.e., lowering Δ) leads to better quality solutions. However, the use of the heuristics can provide solutions of quality comparable to those obtained with the finer resolution ($\Delta = 2$m). On the other hand, as shown in Figure 4b, while computation time rapidly increases with the increase in grid resolution, the approach based on the combined grid heuristics requires much less computational time compared to the case of $\Delta = 2$m, but still providing an equivalently good performance.

5. EVALUATION ON A REAL TESTBED

We validated the solutions proposed by our model using the Indriya testbed [5], located at the School of Computing at the National University of Singapore. It comprises 139 TelosB nodes, deployed in a three floor building.

As a first step, we reconstructed testbed's network topology and made an assessment of the quality of each wireless link in terms of packet reception rate. This was performed by executing at all nodes a basic neighbor discovery protocol, used to collect statistics from periodic broadcasts.

For the experiments, we split the set of nodes into two groups. One group (80 nodes) was selected to become the set of candidate relay positions $\mathcal{R}$ (i.e., potential RNs). The

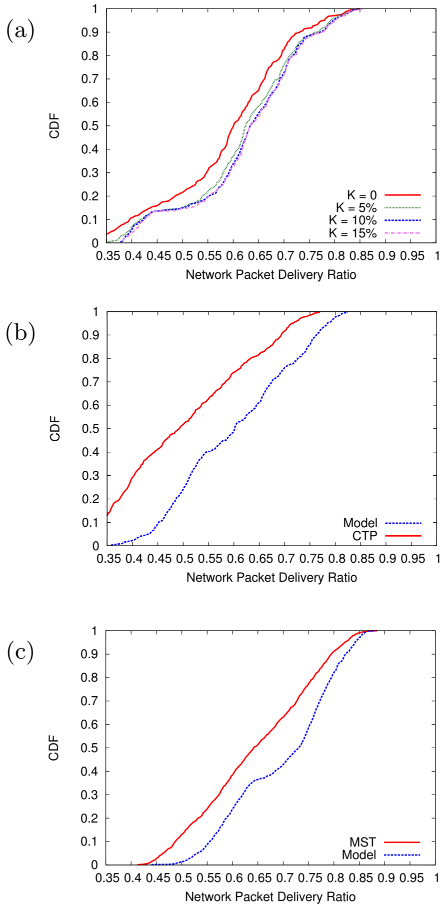


Figure 3: (a) Effect of increasing the percentage K of relay nodes. (b) Comparison between the routing performance of our model vs. that of CTP. (c) Comparison between our model and the MST heuristic for connectivity provisioning.

second group of nodes (59 nodes) was used to define the WSN of reference. Within this WSN, 4 nodes were given the role of BSs. The remaining 55 nodes were given the role of active sensors generating data packets at certain rate and forwarding packets from other sensors towards one of the BSs. All nodes representing a candidate relay position were disabled by default, and do not perform any data transmission unless explicitly selected to become a relay node. Figure 5 shows the considered experimental setup. The network layout shown in the figure does not correspond to the actual positions (which were unknown), and only serves to visualize network topology. Given the reduced problem size, in comparison to the instances considered in the previous section, we were able to compute all solutions in an almost negligible amount of time (less than 2 seconds).

For each experiment, the network packet size was set to 64 bytes. To perform data routing, nodes use pre-defined static routes obtained using the RNP-PE model. No packet acknowledgments nor retransmissions were used, such that a basic best-effort service is provided by the WSN. Nevertheless, we use a simple flow control mechanism, consisting of internal message queues and small random jitters between consecutive transmissions, in order to reduce packet losses. We use the *network delivery ratio*, defined as the total number of packets received at the BSs divided by the number of packets generated by all sensors, as a metric of network throughput. The standard deviation of nodes' delivery ratios, referred to as *network unfairness*, provided a way to evaluate how QoS is distributed throughout the network. For instance, a higher value for network unfairness indicates that some nodes have suffered from larger data losses compared to the average network delivery ratio.

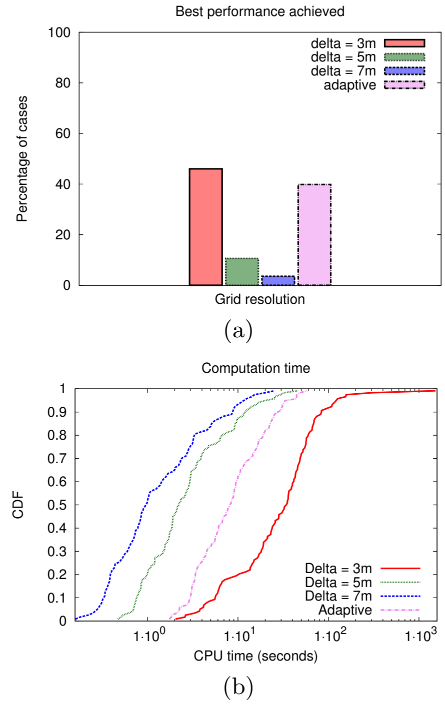
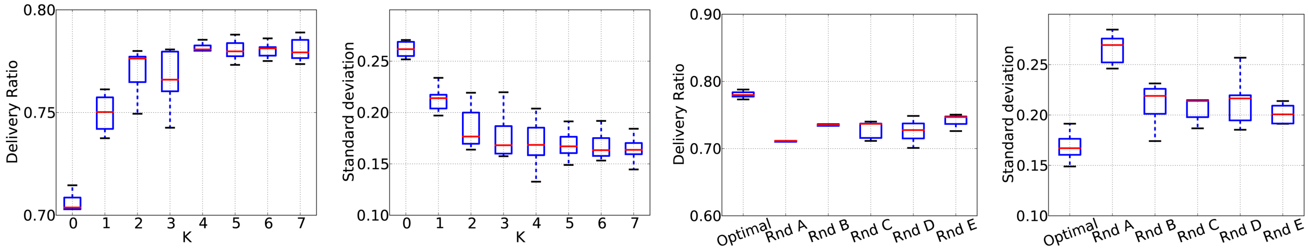


Figure 4: Comparison of grid resolutions (for $K = 15\%$). (a) Solution quality. (b) Computation time.

5.1 Performance enhancement by relay nodes

We use the RNP-PE model to select from $\mathcal{R}$ a limited number $\mathcal{K}$ of nodes to act as RNs. The scenario is equivalent to have $\mathcal{K}$ extra nodes available whose possible deployment is restricted to the positions of $\mathcal{R}$. We increased the value of $\mathcal{K}$ from 0 (i.e., only routing paths are computed), up to the maximum number of RNs that the WSN can exploit (for the considered WSN, this value turned out to be around 7). Nodes generate 4 packets of 64 bytes per second. Model solutions were evaluated in the testbed by enabling the corresponding RNs and setting up the static routes. To account for randomness and temporal variations of the wireless spectrum, we ran 10 experiments, each of them corresponding to 3 minutes, for each single RNP-PE instance.

Figures 6a and 6b show the boxplots of the network delivery ratio and network unfairness for increasing value of $\mathcal{K}$. Results show an effective increase of network throughput (higher network delivery ratio) for the instances using RNs. For larger values of $\mathcal{K}$, the increase on the delivery ratio becomes flat, which can be considered as an indication that the network has reached its peak performance. Similar results can be observed for the network unfairness property. The results also indicates that the RNP-PE model is able to effectively determine the appropriate number of RNs, avoid-



(a) (b) (c) (d)

Figure 6: (a,b) Effect of increasing the number of relays: (a) Network delivery ratio, (b) Network unfairness. (c,d) Comparison between RNP-PE solution and random placement of relays ($K = 5$): (c) Network delivery ratio, (d) Network unfairness.

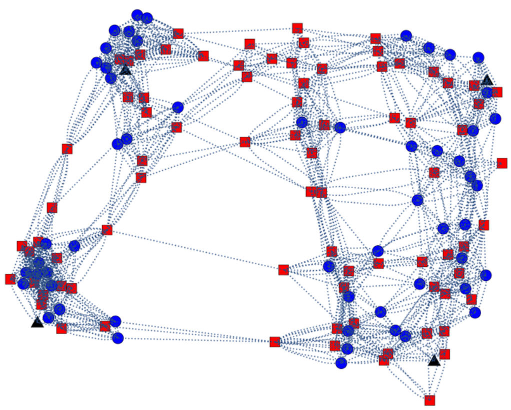


Figure 5: Experimental setup: 4 base stations (triangles), 55 sensor nodes (circles), 80 candidate relay positions (squares).

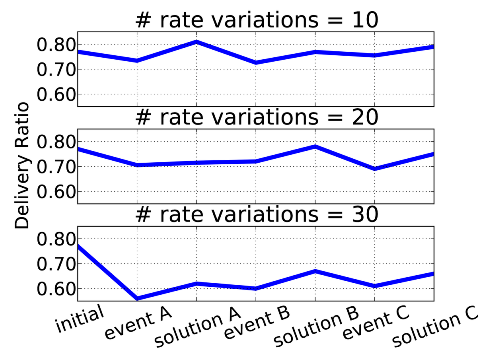


Figure 7: Reactive RNP-PE placement.

ing unnecessary and redundant placements.

In order to evaluate the effectiveness of the RN placements, we evaluated our solution vs. a random selection of RN locations. Considering the case for $\mathcal{K} = 5$, we sampled 5 different sets of RN locations and we computed the associated routing paths. The results, presented in Figures 7a and 7b, show a clear difference in performance between the optimal and the random placements of RNs (for each sample, indicated as Rnd A-E, we performed 10 runs to account for the additional random factors).

5.2 Reactive on-line RNP-PE

In the previous experiments, the model was solved offline. In this section we report about the results for an on-line implementation and use of the RNP-PE model. We consider a reactive setting: node data generation is continuously monitored (e.g., at BSs) to spot significant variations, and when this happens the setup and calculation of a new RNP-PE model is triggered at the BSs. Its output is used for RN repositioning and to determine new routing paths, continually adapting in this way to variations in traffic patterns, (e.g., caused by events such as target detection in surveillance application, or natural events in environmental monitoring).

Initially, all nodes generate data at a rate of 4 packets per second and an initial RNP-PE solution is deployed. In the experiment, we simulate the occurrence of 3 events, namely A, B, and C. At each event, a fraction of the nodes changes their data generation rate (from 2 up to 12 packets per second). Following an event, the network continues for 3 minutes its operations under the previous relay node deployment and routing scheme. At the end of this interval, we measure the current performance and use the RNP-PE model to compute a new relay node deployment and routing scheme, taking into account the new data generation rates. The new solution is then deployed in the network, that is, relay nodes corresponding to the previous placement are disabled, while the ones corresponding to the new solution are enabled, and the new static routing policy is disseminated by the BSs. The dissemination of the static routes is done using a simple controlled flooding scheme. However, more efficient schemes, such as *gossip protocols* could also be used. After a new solution is deployed, the network continues its operation for 3 minutes, when the next event takes place.

Figure 8 shows the variation of performance obtained by the reactive implementation of the RNP-PE model. Three scenarios were evaluated in which we vary the number of nodes that change their data generation rate. The point indicated as *initial*, corresponds to the network delivery ratio before the occurrence of any event. The points indicated as *event A*, *event B*, and *event C* represent the measured performance after the time interval following the occurrence of events. Points indicated as *solution A*, *solution B*, and *solution C* indicate the performance reached after the deployment of the computed solutions in response to each of the events. Results make clear the advantage of using the RNP-PE model to dynamically adjust the position of RNs in a reactive and on-line way in order to tackle the effects of network congestion and routing deficiencies provoked by external events.

6. CONCLUSIONS

In this paper we formalized the problem of relay node placement for performance enhancement in WSNs by defining a linear, mixed integer mathematical programming model. Compared to previous MIP models [7, 22], this is the first comprehensive study focusing both on network performance and connectivity issues, and explicitly including in the model most of the critical aspects of wireless environments.

The model is efficiently solved using a standard solver and outputs both relay node positions and routing paths to base stations. The quality of the computed solutions have been evaluated in simulation and testbed experiments, and validated considering state-of-the-art algorithms for routing and for relay placement for network connectivity.

We also show an on-line application on the real testbed, in which the proposed model is used to continuously adapt the network to variations in the nodes' data generation rate. Optimal solutions, obtained in real-time, enabled the network to maintain a maximal performance by adaptively repositioning relay nodes and recomputing data routes.

In future, we will test our approach using real WSNs and mobile robots, and we will work out its evolution into a fully distributed scheme. On-line estimators will be considered to measure node positions, link costs, and traffic loads.

ACKNOWLEDGMENTS

This research has been partially funded by the Swiss National Science Foundation (SNSF) Sinergia project SWARMIX, project number CRSI22_133059.

7. REFERENCES

[1] I. Akyildiz. Wireless sensor networks: a survey. *Computer Networks*, 38(4):393–422, 2002.

[2] A. Capone and F. Martignon. A multi-commodity flow model for optimal routing in wireless mesh networks. *Journal of Networks*, 2(3):1–5, 2007.

[3] D. Chen. Approximations for steiner trees with minimum number of steiner points. *Theoretical Computer Science*, 262(1-2):83–99, 2001.

[4] X. Cheng, D.-Z. Du, L. Wang, and B. Xu. Relay sensor placement in wireless sensor networks. *Wireless Networks*, 14(3):347–355, 2007.

[5] M. Doddavenkatappa, M. C. Chan, and A. L. Ananda. Indriya: A low-cost, 3d wireless sensor network testbed. In *TRIDENTCOM*, pages 302–316, 2011.

[6] A. Efrat, S. Fekete, P. Gaddehosur, J. Mitchell, and V. Polishchuk. Improved approximation algorithms for relay placement. In *Proc. of the 16th annual European symposium on Algorithms*, pages 356–367, 2008.

[7] S. Ergen and P. Varaiya. Optimal placement of relay nodes for energy efficiency in sensor networks. In *Proc. of ICC'06.*, volume 8, pages 3473–3479, 2006.

[8] E. Falck et al. Balanced data gathering in energy-constrained sensor networks. In *Proc. of ALGOSENSORS*, pages 59–70, 2004.

[9] E. Feo. Optimal relay node placement for throughput enhancement in wireless sensor networks. Master's thesis, Trento, Italy, 2010.

[10] O. Gnawali et al. Collection tree protocol. In *Proc. of SenSys'09*, pages 1–14. ACM Press, 2009.

[11] M. Haenggi and D. Puccinelli. Routing in ad hoc networks: a case for long hops. *IEEE Commun. Mag.*, 43(10):93–101, 2005.

[12] X. Han, X. Cao, E. Lloyd, and C. Shen. Fault-tolerant relay node placement in heterogeneous wireless sensor networks. *IEEE Trans. Mobile Comput.*, 1, 2009.

[13] B. Hao, H. Tang, and G. Xue. Fault-tolerant relay node placement in wireless sensor networks: formulation and approximation. In *Proc. of the IEEE Workshop on High Performance Switching and Routing.*, pages 246–250, 2004.

[14] Y. Hou, Y. Shi, H. Sherali, and S. Midkiff. Prolonging sensor network lifetime with energy provisioning and relay node placement. In *Proc. of IEEE SECON'05*, pages 295 – 304, 2005.

[15] A. Kashyap, S. Khuller, and M. Shayman. Relay placement for higher order connectivity in wireless sensor networks. In *Proc. of IEEE INFOCOM'06*, pages 1–12, 2006.

[16] A. Kashyap and M. Shayman. Relay placement for minimizing congestion in wireless backbone networks. In *Proc. of IEEE WCNC'06*, pages 159–164, 2006.

[17] P. Levis, N. Lee, M. Welsh, and D. Culler. Tossim: Accurate and scalable simulation of entire tinyos applications. In *Proc. of ACM SenSys'03*, pages 126–137, 2003.

[18] S. Li, G. Chen, and W. Ding. Relay node placement in heterogeneous wireless sensor networks with basestations. In *Proc. of CMC*, pages 573–577, 2009.

[19] G. Lin. Steiner tree problem with minimum number of steiner points and bounded edge-length. *Information Processing Letters*, 69(2):53–57, 1999.

[20] E. Lloyd and G. Xue. Relay node placement in wireless sensor networks. *IEEE Trans. Comput.*, 56(1):134–138, 2007.

[21] S. Misra, S. D. Hong, G. Xue, and J. Tang. Constrained relay node placement in wireless sensor networks: Formulation and approximations. *IEEE/ACM Transactions on Networking*, 18(2):434–447, 2010.

[22] M. Patel, R. Chandrasekaran, and S. Venkatesan. Energy efficient sensor, relay and base station placements for coverage, connectivity and routing. In *Proc. of IEEE IPCCC'05*, pages 581–586, 2005.

[23] H.-T. Roh and J.-W. Lee. Joint relay node placement and node scheduling in wireless networks with a relay node with controllable mobility. *Wireless Communications and Mobile Computing*, 12(8):699–712, 2012.

[24] J. Tang, B. Hao, and a. Sen. Relay node placement in large scale wireless sensor networks. *Computer Communications*, 29(4):490–501, 2006.

[25] G. Wang, L. Huang, H. Xu, and J. Li. Relay node placement for maximizing network lifetime in wireless sensor networks. In *Proc. of WiCOM*, pages 1–5, 2008.

[26] G. Wang, L. Huang, H. Xu, and Y. Wang. Reliable relay node placement in wireless sensor network. In *Proc. of ICST*, pages 371–375, 2008.

[27] W. Zhang, G. Xue, and S. Misra. Fault-tolerant relay node placement in wireless sensor networks: Problems and algorithms. In *Proc. of IEEE INFOCOM'07*, pages 1649–1657, 2007.