HEURISTICS FOR DUAL COVERAGE IN WIRELESS SENSOR NETWORKS

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Claudia Sueli Magalhaes Sampaio

In Partial Fulfillment
for the Degree of
MASTER OF SCIENCE

Major Department:
Electrical and Computer Engineering

February 2013

Fargo, North Dakota

North Dakota State University

Graduate School

Title

Heuristics for Dual Coverage in

Wireless Sensor Networks

By

Claudia Sueli Magalhaes Sampaio

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Dr. Roger Green

Chair

Dr. Hong Xiang Li

Dr. Benjamin Braaten

Dr. Vasant Ubhaya

Approved:

April 4, 2013
Date

Dr. Rajendra Katti
Department Chair

ABSTRACT

Nowadays there is an increasing interest in wireless sensor networks due to their usage in a diverse range of applications such as telemetry and smart grids. In such networks, a valid concern is the minimization of the total network cost. To achieve this goal, one objective is to minimize the number of base-stations (nodes) to provide coverage for the sensors. In previous works, it has been usually assumed that base-station coverage already exists over the entire area of interest and the focus is to select cluster-heads in ad-hoc networks. Other works select the base-station among candidates given pre-defined constraints for single coverage. This thesis proposes efficient heuristics to define the number and placement of base-stations for wireless sensor network in a dual coverage scenario. For this purpose, random sensor distributions are generated and comparisons between the proposed heuristics are provided.

ACKNOWLEDGMENTS

Special thanks to my co-advisor, Dr. Li, who gave me the topic, and my advisor, Dr. Green, for all his help and support.

DEDICATION

To my dear parents.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1. INTRODUCTION AND BACKGROUND

Wireless networks have undergone an incredible expansion in the last few decades and have caused a global communications revolution. This phenomenon has occurred not only due to technological innovations, but also due to the simplicity in the implementation of such networks, which makes them a very attractive option for fast and low cost infra-structure deployment.

Initially developed for military purposes, wireless networks are currently present in many fields such as medical, industry and home and are meant to provide all kinds of services. From voice to data, mobile communication to location based services; there is a huge range of services that would be economically impractical, if not technically impossible, in a wired environment. Now, due to wireless technology, such networks are found worldwide.

A specific type of wireless network, the wireless sensor network, has been of increasing importance in the last couple of years due to new applications that are emerging, such as smart grid and telemetry. In such networks, a valid concern is the optimization of the number and placement of base-stations (nodes or access points), as this factor can have a significant impact in the performance and in the overall cost of the network.

1.1. Objective

In this work, we are concerned about the minimization of the number of base-stations in a wireless sensor network, such that each sensor can communicate with at least two base-stations. As there is no polynomial time deterministic approach to solve this problem, we propose heuristics to determine a near-optimal number of base-stations and base-station placement.

We achieve a near-optimal number of base-stations using clustering techniques and we determine dual base-station placement per cluster through a transmission distance minimization criterion. The heuristics we propose are: a) Centroid with Coarse-Grid and Zoom; b) Dual Dominating Independent Set. We provide details of these techniques in Chapter 2, but the main difference between them is that in the first we use the centroids of the positions of the sensors to determine the positions of the base-stations, and in the second we use the number of neighbors (sensors within a pre-establish maximum distance) of the sensors to determine the positions of the base-stations. In the Centroid method, the base-stations can be placed in any position; while in DIS, the base-stations must to be placed in the same position of one of the sensors.

In the next section, we present a summary of related works. In all of them, the wireless coverage already exists over the entire area of interest. Some of the works focus in selecting cluster-heads (sensors selected to communicate to base-stations), others in choosing base-stations among existing candidates, given pre-defined threshold constraints. In our work, the wireless coverage is defined by our algorithms, using the fixed positions of the sensors to determine the base-station placement.

The common feature between our work and the others is the usage of some type of clustering technique. Although clustering elements using the Centroid criteria is not present in any of the related works, the Dominating Independent Set (DIS) appears in [5], [6] and [9]. In [6], DIS is the algorithm with the best performance. In [5], although it is not the one with best performance, it is the one with best time complexity. In [9], as other constraints are considered, DIS is not the appropriate choice for clustering. In our work, our criterion is the minimization of transmission distance, which makes DIS a suitable choice.

This work uses some other ideas extracted from related works. From [4] we extracted the idea of gridding an area, but while there the grid is static, in our work the grid space

dynamically decreases to converge to a near-optimal base-station placement. From [12] we use the criterion of minimization of transmission distance as a constraint for the base-station placement definition.

Regarding contributions, differently from previous works, this work presents a simple and polynomial time algorithm to provide a near-optimal number of base-stations and determine base-station placement in a dual coverage scenario.

1.2. Related Works

The works most closely related to our work are described here. The first two present techniques for selection of base-stations [4] and dual-relay nodes [2] among existing candidates in ordinary wireless networks. The last four works discuss cluster-head selection in ad-hoc wireless sensor networks: [6] introduces the Dominating Independent Set (DIS) algorithm and compares it with the Shifting Strategy algorithm; [5] uses the DIS algorithm combined with constraints for cluster-head selection; [9] presents a weighted recursive clustering algorithm and provides comparison with other algorithms, including the DIS; and [12] presents a two-phase clustering algorithm for extending a sensor's battery life. Next, we give more detailed information about these works.

1.1.1. Selection of Base-Stations

In the first work [4], the focus is on the deployment of picocellula antennas in a wireless local area network (WLAN) and it presents options that vary in complexity and cost. In the first option, named as *user deployment*, it is the user who defines the location of the sites. It is very simple, since the nodes are deployed considering coverage needs and availability of infrastructure without any further analysis, but may result in cell overlap and coverage gaps.

An improvement in the deployment can be achieved without much effort using the *grid installation* approach, where the region to be covered is simply divided into hexagon areas, and a node is placed in each of their centers.

The third option requires the minimization of the contribution of the measurement points with maximum path loss. For the evaluation of the signal quality, a set of M measurement points is selected in the service area and each point is assigned to the node with the minimum path loss.

For this approach, it is required that:

- There exist N candidate nodes.
- There exist M measurement points of path loss.
- The number of final nodes is predefined as K, with $K < N$.

The optimal solution is the set of K nodes out of N which leads to the best value for the objective function (minimum path loss combined with minimum contribution of measurement point with largest path loss). In order to minimize the objective function, instead of using an exhaustive search algorithm, the following heuristic algorithms are proposed:

1- *Pruning*: This algorithm initially starts with a set of N candidate nodes and achieves final K nodes by a process in which at each turn one node out of N is removed, the objective function is computed and the node is included again. After the computation of the objective functions for all the nodes removed at one per turn, the results are compared and the node that when extracted led to the minimum value for the objective function is permanently removed and the number of nodes is decreased by one before the next run.

2- *Neighborhood search*: This algorithm starts with a randomly selected initial set of K nodes. It searches in the neighborhood for a new set of K nodes until it finds one with a lower objective function. Then, it moves to the new set and repeats the process. Otherwise, after reaching a pre-defined number of iterations, the algorithm stops with the last set.

3- *Simulated Annealing*: This random search algorithm gradually decreases the randomness of the search until it reaches convergence. It works by initially defining an arbitrary set of nodes and by randomly modifying this set and comparing their objective functions. If the objective function of the new set is lower than the old one, the current set is kept. Otherwise, the acceptance of the new set depends on a computed probability. The computation of this probability has a degree of randomness that decreases as the number of iterations increases. After a pre-established number of iterations the algorithm stops. The strategy of accepting a new set with worse objective function than the old one according to a computed probability helps to avoid local optimums.

All these algorithms are greedy, because at each step they compute a local optimum. So they may discard a node in a local computation of the objective function that could give the optimal value in a global computation.

1.1.2. Selection of Dual-Relay Nodes

In the second work, Lin et al. in [2] presents a two-phase heuristic for node placement in a dual coverage scenario. The dual coverage is provided through the usage of two *relay stations* (RS) for the communication between the users and a base-station (see Figure 1 extracted from [2]). The users may be fixed or mobile stations and it is assumed that the mobile stations are homogeneously distributed in the cell.

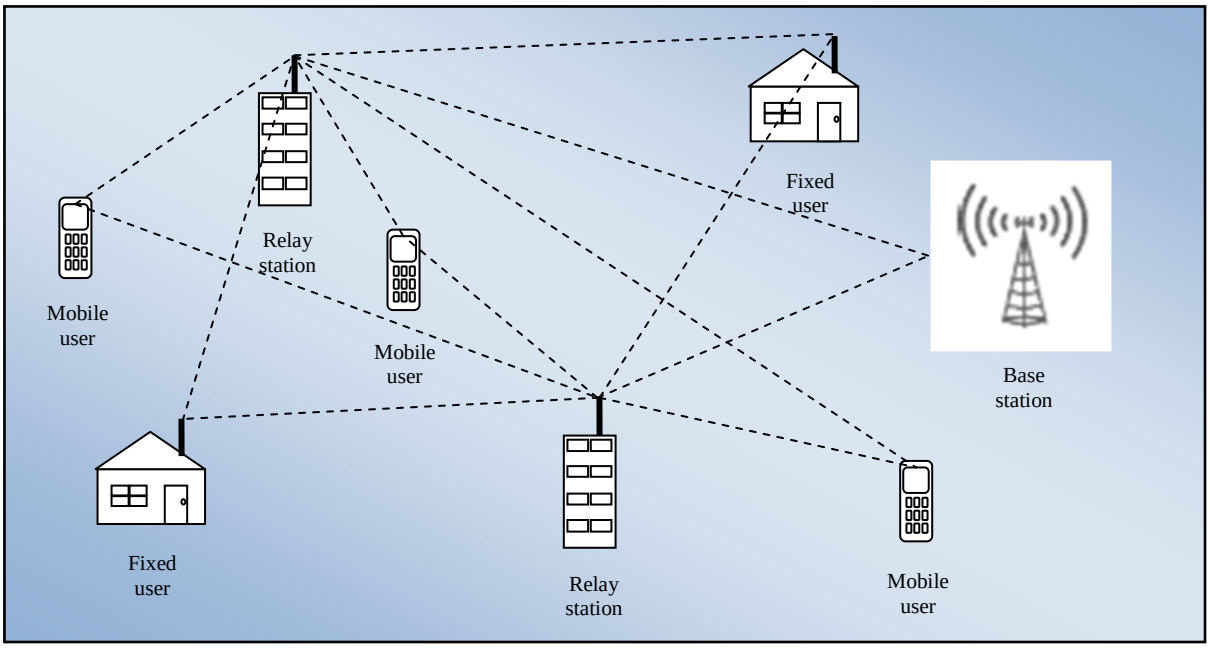

Figure 1. Dual-relay scenario

The optimal placement of a RS is determined by combining the minimal number of RSs, traffic demand per fixed user, and the minimum achievable throughput per bandwidth rate (in bits/s/Hz) per mobile user above a pre-established threshold. The 2-phase algorithm for the node placement consists of:

1- *Pair setting and sorting:* In this phase, the *candidate pairs of RS* (CP) are enumerated. The users are sorted in ascending order of traffic demand and the maximum throughput rate for each CP is computed and stored. Also, the bandwidth for each traffic demand is computed. This phase produces the following lists:

   i. Maximum dual-relay rate per user per CP.

   ii. Bandwidth needed per user's traffic demand per CP.

   iii. CP-pair corresponding to the elements in dual-relay rate list and bandwidth list.

2- *Determining RS pair placement*: In this phase a "local-refine-search" method is applied to find the minimum number of RSs to provide dual-relay for the users and at the same time satisfy the constraints of bandwidth and throughput bandwidth rate,

using the lists obtained in the previous phase. The “local-refine-search” method works by iteratively decreasing the number of CP-pairs and checking if the thresholds for the constraints are violated.

In this approach, it is required that:

- There exists information of demand and throughput bandwidth rate per user.
- There exist candidate pairs of RS (CP), and the number of candidates is bigger than the final number of RS pairs selected.
- There exist pre-established thresholds for the constraints.

This approach is also greedy, so it may provide a sub-optimum solution.

The previous work considered a network where the transmission of data from the users to the base-stations is routed by intermediate points referred as relay stations by a single hop, which means that the transmission of the data from source to destination uses just one intermediate point.

The next examples consider wireless networks where all elements can route data and data routing is dynamically established. Each routing point is called a hop and a network with such property is referred as an ad-hoc network. Ad-hoc networks are multi-hop. Sometimes in the literature, a routing point would be referred to relay-node and a hop would mean the jump between points. The data is relayed until it reaches a cluster-head, where it is forwarded to a base-station. These characteristics are common in wireless sensor networks, where sensors may work as relay-nodes (short transmission range to communicate with other sensors) or cluster-heads (long transmission range to communicate to base-stations).

1.1.3. Selection of Cluster-heads

In this section, we present two works for selection of cluster-heads in a wireless sensor network.

In [6], we have our first example of a wireless ad-hoc network, where the cluster-heads are defined using the Dominating Independent Set (DIS) heuristic. To understand the DIS approach, the wireless ad-hoc network is modeled using graph theory concepts, where the positions of the sensors are represented as vertices V, the connections between each pair of sensors are represented as edges E, and the wireless network is represented as an undirected graph *G(V,E)*. Also, the edges between any pair of vertices are possible only if they are apart from each other within a maximum pre-established distance. In such cases, the vertices are neighbors.

A Dominating Independent Set (DIS) is a sub-set *S* of *V* where at the same time there is no edge between any pair of vertices in S (independence), and a node in *V-S* (node in *V*, but not in *S*) has at least one neighbor in *S* (domination) [6]. Figure 2 shows examples of DIS set (shading vertices) for 1-hop and 2-hops configurations.

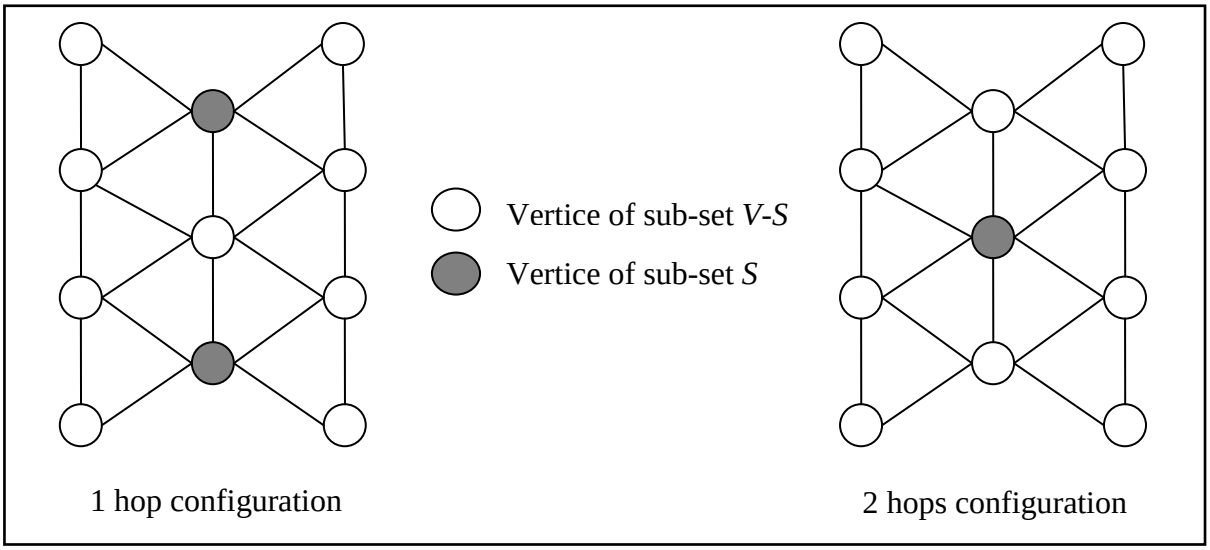

Figure 2. DIS

Assume that:

i. The neighborhood of a vertex *v* is within the range 1 unit of distance.
ii. The graph *G(V,E)* is divided by disjoint clusters, so each vertex belongs just to one cluster.
iii. In each cluster, there is a cluster-head *y* and all the vertices of the cluster are connected to the cluster-head via multi-hop routing.
iv. A cluster has tree depth *R* if the maximum number of hops in a path from a vertex *v* to its cluster-head *y* is at most *R*.

The DIS is an algorithm for multi-hop systems with a tree depth *R*, where the cluster-heads are chosen among the vertices of the network. It selects the minimum Dominating Independent Set by computing the number of neighbors n_v of each vertex in the graph based on the tree depth *R*. At each iteration a cluster-head replaces the vertex with greatest n_v value. This process repeats until all vertices are assigned to a cluster-head within the tree depth *R* range.

Figure 3 (extracted from [6]) shows an execution of DIS, where the circles represent the vertices and the numbers close to the circles represent the number of neighbors n_v of each vertex for a tree depth $R = 5$. Initially, all vertices are uncovered, represented by the gray color. In a) the vertex with greatest n_v (in this case $n_v = 31$) is selected as a cluster-head, represented by the big black circle. In b) the vertices covered by the cluster-head selected previously are in white and a second cluster-heard is selected among the vertices with $n_v = 16$ to cover the remaining vertices in gray.

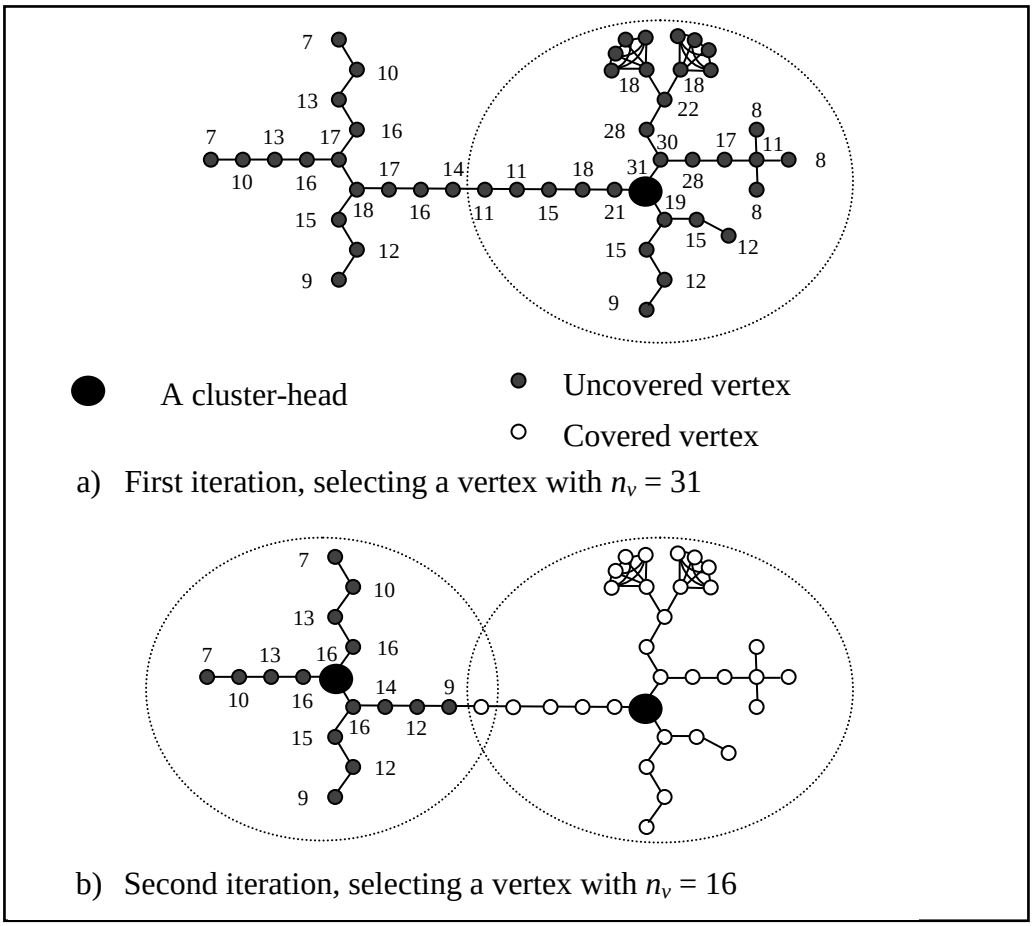

Figure 3. Example of execution of DIS

The next step eliminates any eventual overlapping of clusters to create disjoint clusters by assigning each vertex to its closest cluster-head. If the cluster-heads are at the same hop-distance of the vertex, the vertex is assigned to the cluster with smaller weight, or number of vertices. In Figure 3 b) there is one vertex in the overlapping region, shown in the middle of the two dashed circles. Although this vertex was initially assigned to the cluster in the right, it will be reassigned to the cluster in the left. This is because this vertex is at the same hop-distance of both cluster-heads, but the cluster-head in the left has a smaller weight.

After the creation of disjoint clusters, the algorithm checks the weighting and relay-load requirements. If a cluster-head has more vertices than it can deal with or a

vertex is relaying more data than it can support, more cluster-heads are added in the system.

In [6], we have:

- The algorithm continues until all nodes are connected.
- Cluster-heads are chosen among the existing nodes.
- Each node must have at least one neighbor.

In [6], a method called the Shifting Strategy is also described. The Shifting Strategy leads to better performance (smaller number of cluster-heads), but worse running than the Dominating Independent Set (DIS). The details of the Shifting Strategy are omitted here as the better running time of DIS makes it an appropriate choice for comparison purposes.

Another work about selection of cluster-heads in ad-hoc networks is presented by Chehri et al. in [5]. This work considers a static hierarchical ad-hoc wireless network composed of tiny and cluster-head sensors. It is static because once a path of communication is defined, it cannot be dynamically modified. It is hierarchical because the tiny sensors are assumed to have shorter transmission range and lower hop capabilities than the cluster-heads. After reaching a cluster-head the data is routed to a wired backbone. As the tiny sensors are cheaper than the cluster-heads, the minimization of the number of cluster-heads contributes to the minimization of the overall network cost. Therefore, the goal is to find the best selection for the cluster-heads among the sensors positions such that the following constraints are respected:

i. Connectivity: all elements (tiny sensors and cluster-heads) must be connected
ii. Network cost: the number of cluster-heads must be equal or less than a maximum pre-established value D_{max}.

iii. Quality of service (QoS): the tiny sensors must reach a cluster-head in at most h_{max} hops in order to be compliant with allowed time delay and bit error rate for acceptable transmission.

The proposed approach used the Greedy Dominating Independent Set (DIS) heuristic presented in the previous work in [6], with the following differences:

- DIS is used just to define the clusters. Once a cluster is defined, the cluster-head is placed in the center of the cluster in such a way that it is also within the transmission range of at least one of the previously defined cluster-heads.
- The maximum number of cluster-heads is pre-defined D_{max}. In case D_{max} is not big enough, the algorithm stops before the coverage is assured to all nodes.
- A tiny sensor may not be within the transmission range of another tiny sensor. In this case, the isolated tiny-sensor must become a cluster-head.

1.1.4. Selection of Gateways

The next work is about a weighted recursive clustering algorithm for gateway placement in wireless mesh networks presented by Aoun et al. in [9]. The gateways here have the same characteristics of the cluster-heads mentioned before, except that, instead of being used to transmit data to and from a base-station, the gateways are directly connected to fixed network.

The basic recursive version of this algorithm consists in selecting the dominating set of a graph to be the gateway nodes of the network. The selection of such nodes is done by a greedy approach and recursion, respecting the following quality of service (QoS) requirements: cluster radius (number of hops allowed from a node in the cluster to the gateway), relay load (traffic supported by a relay node), and cluster size (number of nodes in a cluster). At each iteration, the set of nodes selected in the previous iteration is used to select

the new set until the threshold of the cluster radius is reached. The weighted version adds a weight per node which indicates the sum of nodes covered by a node, with higher weight for closer nodes (small number of hops) and it helps to compute the cluster size requirement.

The weighted recursive algorithm is compared to the basic recursive algorithm, the iterative greedy given in [6], and a third option, the augment algorithm. The weighted recursive algorithm presents the best performance, giving the smaller number of gateways. The reason for this difference in performance is because in the weighted and basic recursive algorithm, the clusters are formed only when all the requirements are fulfilled. In the other algorithms, first the clusters are defined using the radius constraint, and only after that the other QoS requirements are checked. So, if the QoS requirements are not respected, it triggers the creation of sub-clusters.

1.1.5. Two-phase Clustering for Battery Life Extension

The last reviewed work related to wireless ad-hoc networks is given by Choi, Shah and Das in [12]. They developed a two-phase clustering algorithm for wireless ad hoc networks to promote data aggregation and battery life extension of the sensors when uniformly distributed in a region.

The goal of the first phase is cluster formation, with the selection of one cluster-head and the members of each cluster by a process of advertisement broadcast from cluster-head candidates and join-request messages from neighbors. Every member stores the messages received in order to identify which neighbor is the cluster-head and which neighbors belong to the same cluster.

Once the clusters are defined with their respective cluster-heads, the second phase, to define connections among members of the cluster, takes place. In this phase, each cluster-

head selects the farthest member of the cluster, which has to identify its closest neighbor in the cluster to relay data. For some applications, close members are more likely to have redundant data. If this is the case, data aggregation may also be performed. The process continues with each member of the cluster selecting the closest neighbor that has not been selected yet to relay data until all the members in the cluster are selected or the maximum number of hops allowed is achieved. In both situations the last member selected relays data to the cluster-head. If there are members which are still not relaying data, the cluster-head selects the farthest member among them and the process repeats.

Cluster-head rotation, or reselection of a cluster-head, can also be performed inside each cluster. The cluster-head rotation is triggered by the current cluster-head as soon as a pre-established threshold of its battery level is reached. The current cluster-head selects a new cluster-head with the highest battery level among the members of its cluster.

Simulation results shows a significant reduction of the transmission distance and, consequently, of the power consumed in the network when the second phase is compared to the first phase (member relaying data to closest neighbor versus member relaying data directly to cluster-head).

1.3. Final Comments about Related Works

Many of the previous works considered ad-hoc networks and the multi-hop property. This property can considerably decrease the total cost of the networks, as relay stations are cheaper than base-stations. On the other hand, a multi-hop network is less robust and a failure in a relay station can entirely interrupt the communication in one path. For some applications this may not be acceptable. In the next chapter, we address the base-station coverage problem for wireless sensor networks to find near-optimal number of nodes and node placement in a dual coverage scenario.

CHAPTER 2. PROBLEM FORMULATION AND SYSTEM MODEL

In this work, we are focusing on optimizing the number of base-stations and determining base-station placement in a wireless sensor network, such that each sensor can communicate with at least two base-stations.

The optimal solution for this problem requires an exhaustive search among all possible base-station locations. An alternative approach is to develop polynomial-time heuristics to determine a near-optimal solution for the problem. In this chapter, we propose two heuristics to solve our problem: a) Centroid for clustering and Coarse-Grid and Zoom for base-station placement; b) Dominating Independent Set (DIS) for clustering and Dual DIS for base-station placement.

This chapter is organized as follows: in Sec. 2.1 we introduce the problem statement and related assumptions. In Sec. 2.2 we present the initial considerations, and set the lower and upper bound of our problem. In Sec. 2.3 we present the system model and discuss the general approach of the problem. In Sec. 2.4 we propose our first heuristic, and in Sec. 2.5 we detail our second heuristic.

2.1. Problem Statement

Given a set of wireless sensors distributed in a region, the optimization problem is to find the minimum number of base stations and their placement such that each sensor can communicate with at least two base-stations. For this problem, we assume:

i. locations of the sensors are fixed and given,
ii. coverage of a base-station is known and equal to radius *R*,
iii. each sensor must be able to communicate directly with at least two base-stations. Isolated sensors must communicate with at least one base-station.

iv. each base-station must be placed in a distinct location (geographic diversity). Base-stations serving isolated sensors must be placed in the same location of the sensors.

The first assumption restricts the analysis to a static network scenario as the location of the sensors is fixed. Since the location of the sensors is also given, we determine the position of the base-stations based in the position of the sensors.

For the second assumption, according to Friis Equation [16], the coverage radius of a base-station, considering the electromagnetic signal propagating in free space and the same attenuation in all directions, is equal to:

$$R = \frac{\lambda}{4\pi}\sqrt{\frac{P_{t,max}G_tG_r}{P_{r,max}}} \quad (1)$$

Where λ is the wavelength of the electromagnetic signal, and $P_{t,max}$ is the maximum power of the transmitting antenna, $P_{r,max}$ is the maximum power of the receiving antenna, G_t is the gain of the transmitting antenna, and G_r is the gain of the receiving antenna. In this work, we assume that all base-stations have the same coverage radius.

The third assumption is to establish the dual coverage for the sensors. The single coverage is accepted only if the cluster has just one sensor. Also, the fact that the sensors must be able to communicate directly to the base-stations means that any sensor can be elected as a cluster-head. Thus, it is possible to perform a cluster-head rotation with all sensors that communicate directly with the base-station, in order to help extend the sensors' battery life.

The last assumption requires that the dual coverage is provided with geographic diversity. Besides redundancy, geographic diversity brings another advantage if pre-established constraints need to be respected: the position of the base-stations can be chosen

such that the constraints are minimized. In our work, we use the power minimization criterion as a constraint to determine base-station placement.

2.2. Initial Considerations

Base-station placement for a set of sensors is dependent of density and distribution. For a *low density and scattered sensor distribution*, equal or less than one sensor per base-station coverage area, the sensors will be isolated from each other. In such case, there is not much optimization possible for the number of base-stations to cover them. This implies the worst number of base-stations compared to the number of sensors, which is equal to two times the number of sensors in the dual coverage scenario. Figure 4 shows a scenario with 10 isolated sensors and their respective base-station placement areas, indicated by the shading areas. In a dual coverage scenario, there are two base-stations per placement area. Note that, as the placement areas do not overlap, even if a base-station is placed in the border of a placement area, it can only cover the sensor of that placement area.

For a medium to high density and uniform sensor distribution, an effective approach is to use a method known as grid installation [4]. As mentioned before, in this method, the coverage of the region can be considered rather than the coverage of the sensors. So, the region is simply divided into hexagon areas equivalent to the coverage area of the base-station. One base-station or two base-stations are placed in each area for single or dual coverage respectively. Figure 5 shows the grid coverage, which requires 46 coverage areas of unit coverage radius for a region of size 10 x 10. In this case, the position of the sensors is not relevant and there is some waste of resources, since overlap, represented by the dark shading, is required to cover the whole area.

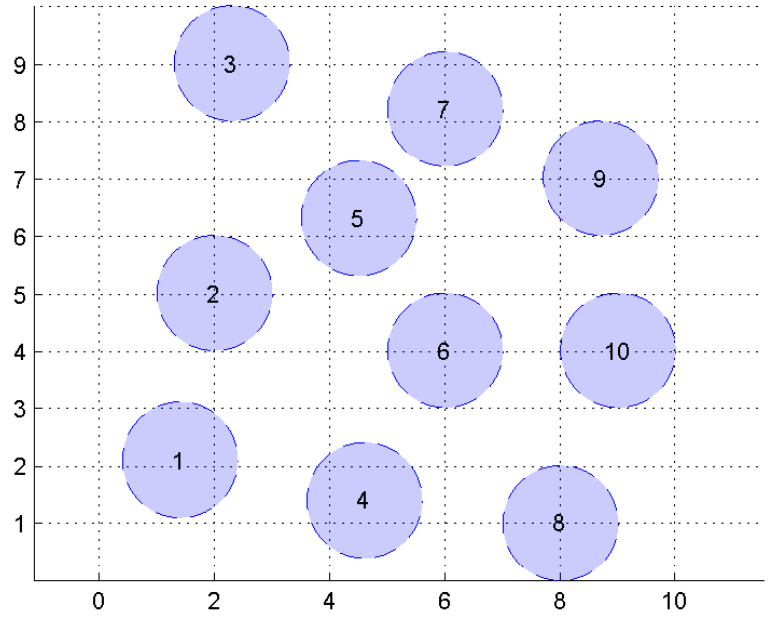

Figure 4. Base-station placement areas for low density sensor distribution

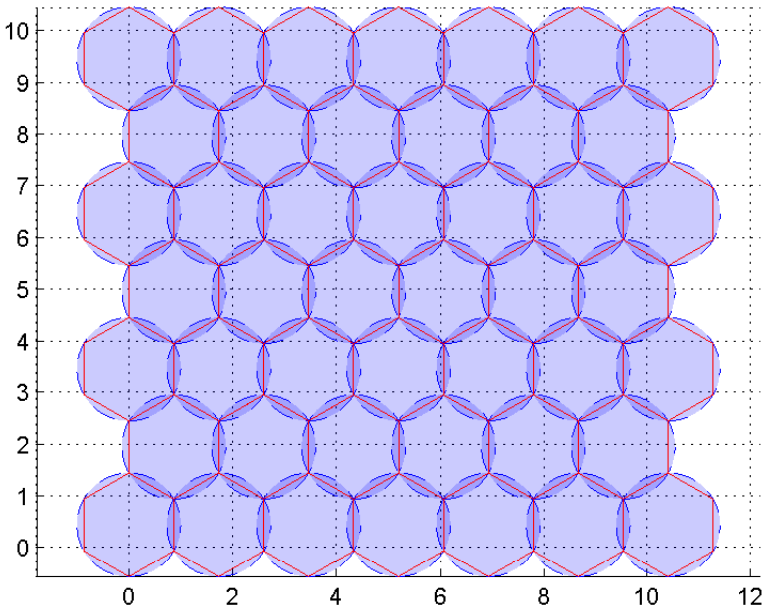

Figure 5. Base-station placement areas for grid installation

2.3. Methodology and System Model

Finding the number of base-stations and the base-station placement in the previous two sensor distributions scenarios are simple. However, if the distribution of sensors is *non uniform*, the optimization of the number of base-stations is not trivial and requires clustering methods. In the lack of a deterministic approach, an exhaustive search among all possible clustering combinations produces an optimal solution, but leads to unrealistic running times. Such problems are classified as non-deterministic polynomial-time hard (NP Hard) [15].

The non-deterministic polynomial-time hard problems are usually optimization problems derived from the non-deterministic polynomial-time complete (NP Complete) problems, which are decisional problems, the most famous being the "Traveling Salesperson". In the general "Traveling Salesperson" problem, given a weighted graph (vertices connected to each other through edges of different lengths), the objective is to find a tour that starts and ends in the same vertex and visits all the other vertices of the graph exactly once. The decisional "Traveling Salesperson" problem tries to answers if, given a number *d*, there exists a tour with total weight (sum of lengths of edges of the tour) no greater than *d*; while the optimization "Traveling Salesperson" problem searches for the tour with minimum total weight.

As our problem is NP-Hard, an alternative approach to achieve a polynomial time solution is to develop heuristics. The heuristics may lead to sub-optimal solutions, and a comparison between methods is required in order to choose the one that leads to the best results. In our work, we propose two heuristics to provide near-optimal solutions of the number of base-stations and base-station placement to provide dual coverage for a wireless sensor distribution.

First, we model our problem considering sensors, cluster-heads and base-stations as vertices *V*, the connections between vertices as edges *E*. The network formed by sensors and base-stations is modeled as an undirected graph *G(V,E)*. Edges are possible only if are equal or less than the coverage radius: $E \leq R$. Other parameters for our system model are:

$S = \{s_1, \ldots, s_N\}$: set of sensors, where $|S| = N$: cardinality of *S*,

$C = \{c_1, \ldots, c_M\}$: set of clusters, where $|C| = M$: cardinality of *C*,

$B = \{b_1, \ldots, b_L\}$: set of base-stations, where $|B| = L$: cardinality of *B*,

$U = \{u_1, \ldots, u_L\}$: set of cluster-heads, where $|U| = L$: cardinality of *U*, and

(x_i, y_i): position of the vertex *i* (sensor, cluster-head or base-station) in the two-dimensional plane.

In the next sections we describe our heuristics: a) Centroid with Coarse-Grid and Zoom; b) Dual Dominating Independent Set (DIS).

2.4. Centroid with Coarse-Grid and Zoom

In this method, we do not define cluster-heads. Also, as one of the assumptions is that each sensor can communicate directly with at least two base-stations, we initially consider edges just between sensors and base-stations, such that: $E \leq R$.

The method has in two phases: cluster formation, with the determination of the minimum number of clusters and the corresponding base-station placement areas; and dual coverage, with the determination of near-optimal base-station placement.

2.4.1. Cluster Formation

In this work, the identification of clusters of sensors is based on a proximity measure [1]. The Euclidean distance between a pair of sensors can be considered as a proximity measure, and is given by:

$$d(s_i, s_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (2)$$

Where (x_i, y_i) *and* (x_j, y_j) are the coordinates of sensors s_i and s_j respectively in a two-dimensional plane.

Therefore, the basic cluster unit is defined as a *set of sensors with Euclidean distance* $d(s_i, s_j)$ *between each sensor* s_i *and each other sensor* s_j *in the set, with* $i \neq j$*, equal or less than two times the radius R of the base-station.*

$$d(s_i, s_j) \leq 2R \quad (3)$$

If $d(s_i, s_j) = 2R$ for any sensor s_i and sensor s_j inside the cluster, the base-station can only be placed in a single point which corresponds to the center point of $d(s_i, s_j)$. In real cases, this is very restrictive because it may be difficult to place the antenna in that specific location. Also, a single point does not satisfy the dual coverage scenario with geographic diversity, because the two base-stations would be placed in the same location. So, if necessary, a constraint about the minimum distance between the two antennas of the cluster or the minimum area for the base-station placement can be added.

The objective function for cluster formation minimizes the number of base-stations through the minimization of the number of clusters:

$$\min[|C|] \quad (4)$$

Subject to:

$$d(s_i, s_j) \leq 2R, \forall\, s_i, s_j \in c_k \tag{5}$$

$$E(s_i, b_j) < R, \forall\, s_i, b_j \in c_k \tag{6}$$

$$|B_{c_k}| = \begin{cases} 1 & \text{if } |S_{c_k}| = 1 \\ 2 & \text{if } |S_{c_k}| > 1 \end{cases} \tag{7}$$

To accomplish a near-minimum number of clusters, the proposed algorithm is based in the centroid of the sensors. The centroid of the set of sensors is computed through the arithmetic mean, given by:

$$\mu_{c_k} = \left(\frac{\sum_{i=1}^{N_{c_k}} x_i}{N_{c_k}}, \frac{\sum_{i=1}^{N_{c_k}} y_i}{N_{c_k}} \right), \tag{8}$$

where N_{c_k} is the total number of sensors in the cluster.

Initially, the algorithm considers all the sensors of the set in the same cluster and computes the centroid for all of them, so $N_{c_k} = N$. In the next step, the algorithm computes the distance between the farthest sensor and the centroid as if there is a base-station in that position. If this distance is greater than the threshold *R* (see equation (6)), the farthest sensor is discarded from the cluster and a new computation of the centroid is performed considering the remaining sensors. Otherwise, a cluster is determined. Once the cluster is determined, if there are discarded sensors, they are assigned to another cluster and the process of computing the related centroid and checking the distance of the centroid and the farthest sensor of the remaining set is repeated until all the sensors are assigned to a cluster. In this method, the closer the sensors are to each other, the more likely they will be assigned to the same cluster.

The determination of clusters using the centroid method is greedy; it may discard a point in a cluster for a local computation that in a later step may fit in the cluster. To diminish this effect, a refinement can be applied. Once the clusters are defined, the distance of the centroid of each cluster and the farthest point of each other cluster can be computed. If the distance is less than the threshold *R*, the clusters are merged. The application of the centroid algorithm without and with the refinement is shown in Figure 6 and Figure 7 respectively. In both figures, the clusters are showed as dashed circles. The positions of the sensors are indicated by their cluster numbers, and the edges are the solid lines connecting the sensors to the centroid of each cluster. In Figure 6 there are 14 clusters. It can be noticed that the three sensors in cluster 12 (in the bottom left on the figure) could have been assigned to cluster 5 as they are inside both of the clusters. This is because, during the computation process, the centroid in cluster 5 was shifted closer to sensors in cluster 12, as sensors in other neighboring clusters were discarded in each turn. Refinement solves this situation as seen in Figure 7, where cluster 12 is merged with cluster 5 and the total number of clusters decreases to 13.

The refinement improves the performance of the algorithm at the cost of a higher computational time complexity. Table 1 shows the time complexity comparison between the Centroid clustering without and with refinement. In the Centroid clustering without refinement, the algorithm has to compute the centroid and the farthest sensor for every cluster, giving a time complexity of $\mathcal{O}(N^2)$. In the Centroid clustering with refinement, besides the computation of the centroid and the farthest sensor for every cluster, the algorithm has to check each cluster can be merged with the others, giving a time complexity of $\mathcal{O}(N^3)$.

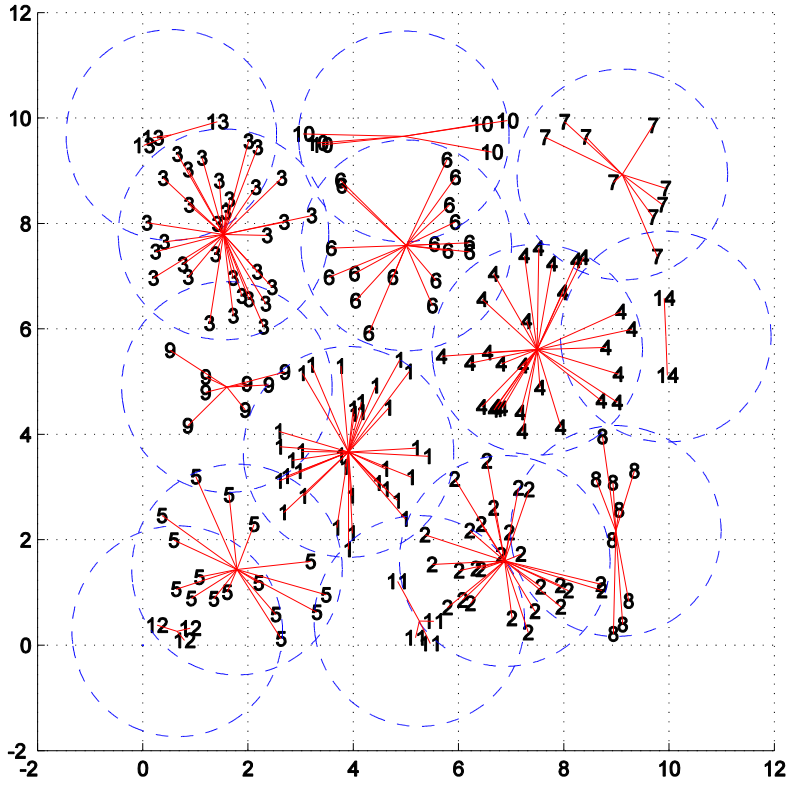

Figure 6. Clustering without refinement

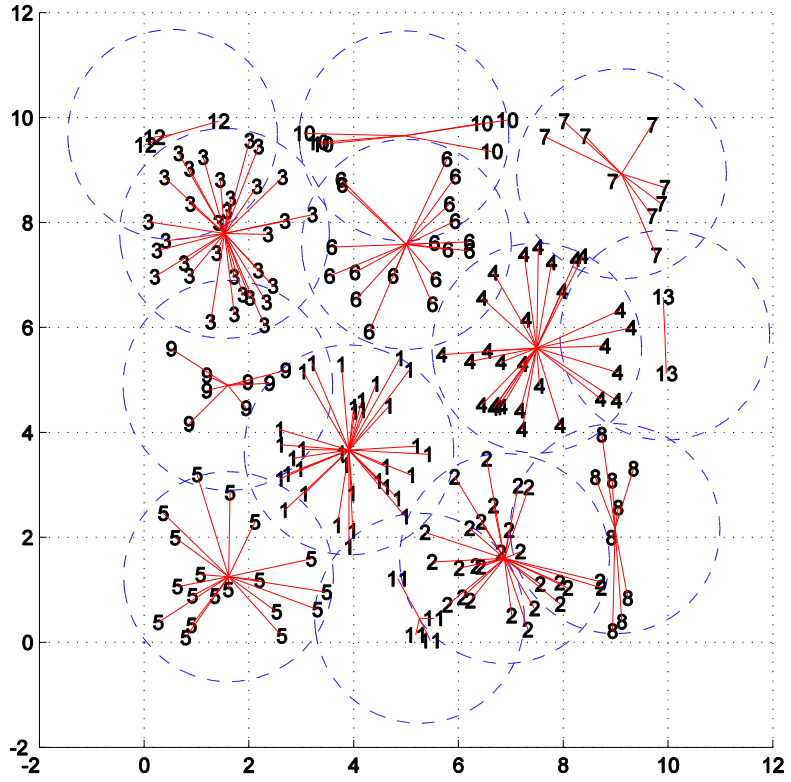

Figure 7. Clustering with refinement

Table 1. Time complexity comparison

Centroid clustering without refinement	Centroid clustering with refinement
$\mathcal{O}(N^2)$	$\mathcal{O}(N^3)$

After clustering, the next step is finding the base-station placement area inside each cluster, which is determined by the intersection of the coverage areas of all the sensors belonging to the cluster. In order to find the intersection area, we apply the bisection method in each cluster. The purpose is to determine the border of the placement area. To do so, we need to define the number of border points and the number of iterations. The number of border points is used to set radial lines starting from the centroid, while the number of iterations defines the number of times we bisect each of the radial lines. Initially the bisection of each radial line occurs between the centroid and the border point. Then, we check if all the sensors in the cluster are covered if a base-station is placed in the bisection point. If all sensors are covered, we bisect the sub-interval from that point to the radius *R*, otherwise we bisect the sub-interval from the centroid to that point. We keep bisecting into smaller sub-intervals until we reach the number of iterations. In our work, we use 256 border points and 10 iterations. This means that, in the worst case, for a coverage radius of 1 unit length, the distance between border points is 0.0245 (=2πR/256).

As the bisection method is an approximation, the borders points may present some variations to the real position. The application of the convex hull method assures that only the most external border points are considered to define the border line of the area.

Figure 8 shows the application of bisection method and convex hull for a simple example with 8 radial lines and 4 iterations. In this example, there is just one cluster, represented by the dashed circle, with 3 sensors, represented by small circles. The centroid is indicated by a square. The numbers shows the position of the iterations in each of the radial lines. The patched area is the result of the application of convex hull.

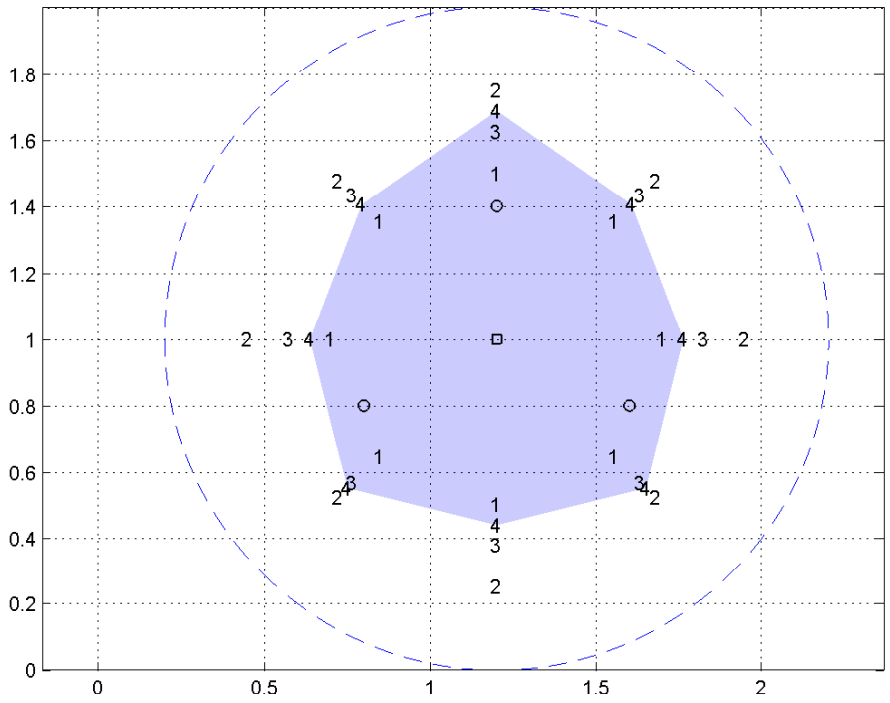

Figure 8. Bisection iterations and convexhull

Figure 9 shows a more realistic example of computing base-station placement areas using the bisection and convex hull methods. A set of one hundred sensors is divided in twenty six clusters and the positions of the sensors are marked with the number of the cluster to which they belong. The dashed circles represent the coverage area of a base-station with radius *R* when placed in the centroid of the sensors belonging to the cluster. The sensors connect to the centroid through solid lines. The

patched area in each cluster is the placement area for the base-stations. This area corresponds to the intersection of areas that would result if base-stations were placed in the position of the sensors of the cluster. So, a base-station placed in any position inside the placement area will cover all the sensors of the cluster.

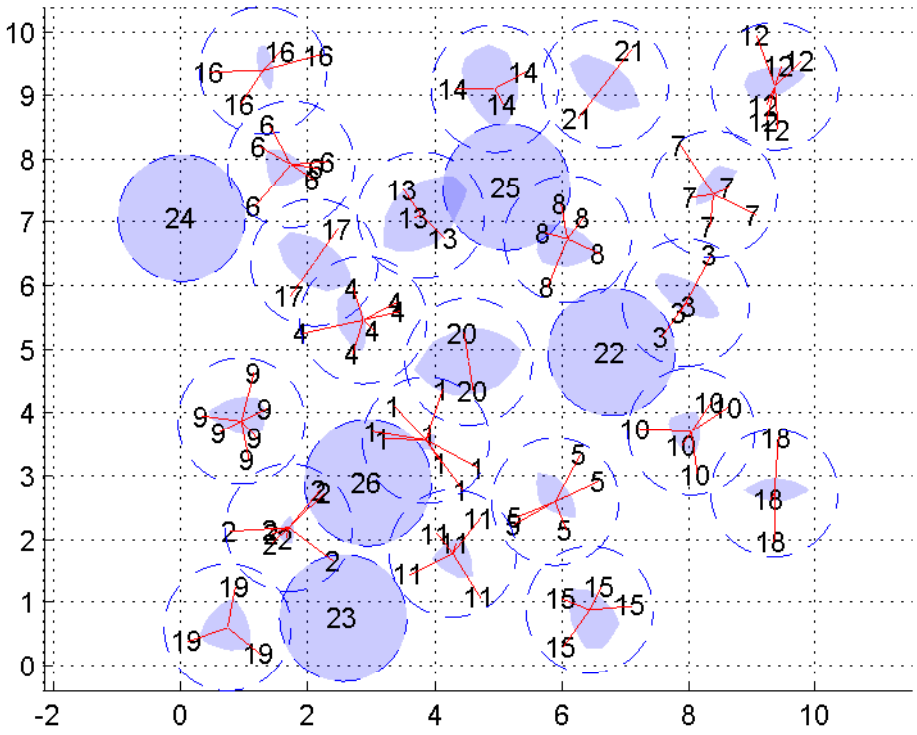

Figure 9. Clusters and base-station placement areas

The pseudo-code to find clusters and base-station placement area is given in Table 2.

Table 2. Pseudo-code for clustering

Input: Number of sensors *N*, sensor positions (x_i, y_i)

1. Initialization: $k = 1$
2. Assign all s_i to c_k
3. Compute centroid μ_{c_k} for c_k
4. For all $s_i \in c_k$ compute $E(s_i, \mu_{c_k})$
5. **If** $\max(E(s_i, \mu_{c_k})) \geq R$
    6. Discard s_i with $\max(E(s_i, \mu_{c_k}))$ from c_k
    7. Go to 3
8. **Else**
    9. Save c_k
10. **End if**
11. **If** discarded $s_i \neq 0$
    12. $k = k + 1$
    13. Assign all discarded s_i to c_k
    14. Go to 3
15. **End if**
16. **For** $k = 1$ to M
    17. Apply bisection and convex hull in c_k
    18. Save base station placement area A_{c_k}
19. **End for**

Output: Clusters c_k, total number of clusters *M*, base station placement areas A_{c_k}

2.4.2. Base-station Placement

In this phase, a near-optimal base-station placement inside the placement area is determined. The criterion used is the minimization of power consumption for extension of sensors' battery life. We grid the placement area of each cluster and choose two of the grid points that lead to the minimum distance to the sensors. These grid points are selected as the position of the base-stations. The gridding process can be done using a fine grid, with a large number of grid points with a small distance among them, or a coarse grid and zoom that considers a sparser grid (smaller number of grid points) and a zoom factor. In the last option, once the two grid points of the coarse grid are chosen, the method zooms the area around these two grid points for a re-gridding process, dividing the grid space by a zoom factor, until a pre-defined minimum grid space is reached. The grid and zoom method improves efficiency as can be seen from Table 3.

Table 3. Comparison between fine grid and coarse grid and zoom methods

Fine grid	**Coarse grid and zoom**
Given N_{c_k} sensors in a cluster and G_{c_k} possible locations (L' grid points) for 2 BSs. Number of possibilities: $N_{c_k} \times \binom{G_{c_k}}{2} = \frac{\left(N_{c_k} \times G_{c_k} \times \left(G_{c_k} - 1\right)\right)}{2}$	Given N_{c_k} sensors in a cluster and L' possible locations (G_{c_k} grid points) for 2 BSs and z zoom factor. Number of possibilities: $z \times \left(N_{c_k} \times \binom{\frac{G_{c_k}}{z}}{2}\right) = \frac{\left(N_{c_k} \times G_{c_k}\left(\frac{G_{c_k}}{z} - 1\right)\right)}{2}$

As an example consider a cluster with N_{c_k} = 8 sensors, G_{c_k} = 100 grid points and z = 10 zoom factor. Then:

Number of possibilities with fine grid = 39,600.

Number of possibilities with coarse grid and zoom = 3,600 (in this case, ~z times smaller).

In all simulations of this work, we use a zoom factor z = 10, and the re-grid loops until the grid space reaches a pre-defined minimum grid space e. Figure 10 shows an example of dual base-station placement, i.e., two base-stations defined per cluster using the Coarse-Grid and Zoom method with an extension of battery life criterion. In this example, there are one hundred sensors and twenty seven clusters. The base-stations are represented by squares, the positions of sensors are marked with the number of the clusters they belong, and solid lines connect the sensors to the closest base-stations of the cluster. A detail of the dual base-station placement delimited by the big rectangle in Figure 10 is given in Figure 11.

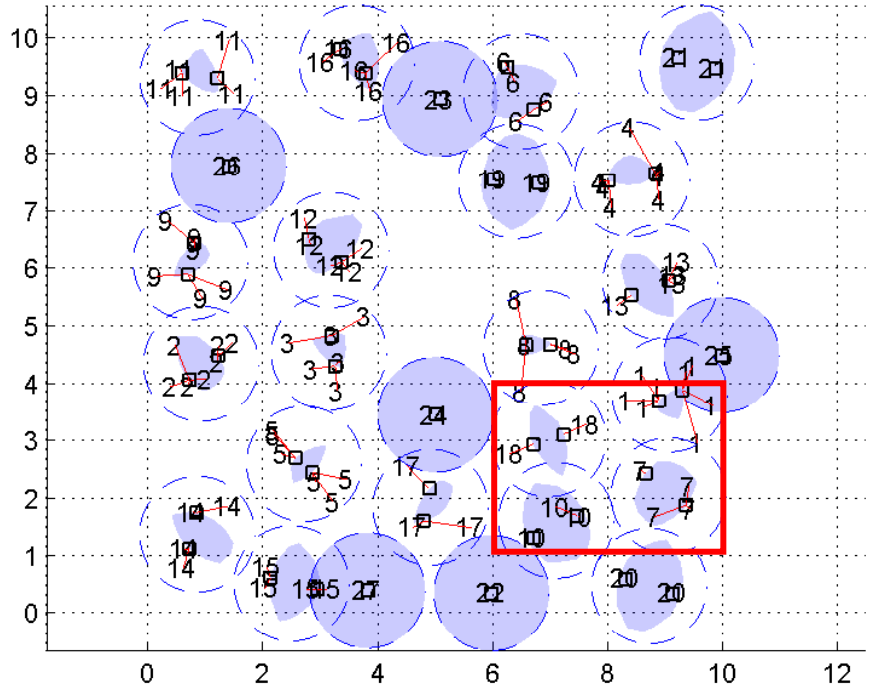

Figure 10. Dual base-station placement

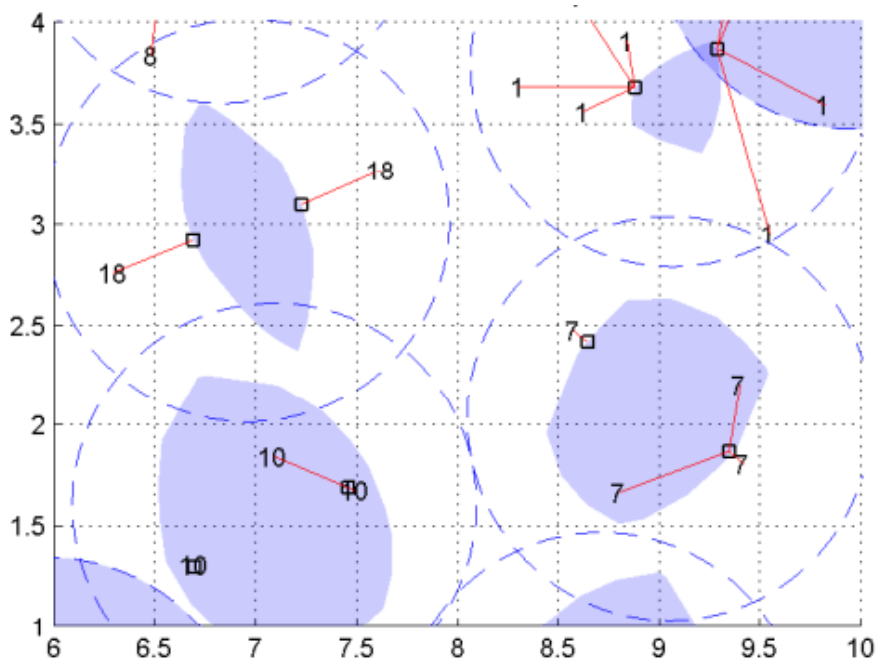

Figure 11. Detail of dual base-station placement

The pseudo-code for the base-station placement in one cluster with more than one sensor using the Coarse-Grid and Zoom is given in Table 4.

Table 4. Pseudo-code for base-station placement

Input: Sensors s_i of cluster c_k, sensor positions (x_i, y_i), base station placement areas A_{c_k}, zoom factor *z*, minimum grid space *e*.

1. Grid A_{c_k}.
2. Find grid points $g_{c_{k1}}$, $g_{c_{k2}}$, and divide S_{c_k} into subsets S'_{c_k}, S''_{c_k} | $\sum_i (d(s'_i, g_{c_{k1}}) +$ *dsi'',gck2)* is minimum.
3. **Zoom: If** $grid\ space > e$
    4. $A_{c_{k1}} = g_{c_{k1}} \pm e$, $A_{c_{k2}} = g_{c_{k2}} \pm e$, and $e = e/z$.
    5. Grid $A_{c_{k1}}$, $A_{c_{k2}}$.
    6. Find new grid points $g_{c_{k1}}$, g_{k2} | $\sum_i (d(s'_i, g_{c_{k1}}) + d(s''_i g_{c_{k2}}))$ is minimum.
    7. Go to 3.
8. **Else**
    9. Define $b_j = g_{c_{k1}}, b_{j+1} = g_{c_{k2}}$, store (x_j, y_j), (x_{j+1}, y_{j+1}), $j = j + 2$.
10. **End if**

Output: Base-stations b_j, base-station placement locations (x_j, y_j).

2.5. Dual DIS

The Dual DIS uses the Dominating Independent Set (DIS) algorithm, already explained in section 1.1.3, but instead of one cluster-head, we select two cluster-heads per cluster. The idea is to place base-stations in the same position of the cluster-heads selected to provide dual coverage for the sensors in the cluster.

In this method, edges are allowed between sensors as long as an edge is no greater than the coverage radius: $E \leq R$. Also, we restrict the number of hops h to be less than or equal to a pre-establish $h_{\max}$: $h \leq h_{\max}$. As one of the assumption requires that each sensor must be able to communicate directly with the base-stations, in our work we have $h_{\max} = 1$.

This algorithm has three phases: cluster-head selection, cluster formation, and dual cluster-head selection.

2.5.1. Cluster-head Selection

In the cluster-head selection phase, some sensors are selected as cluster-heads. The criterion to select sensors as cluster-heads considers the minimum set of sensors with the greatest number of neighbors, such that any other sensor can reach one of the sensors from the minimum set within at most $h_{\max}$. Consider that each cluster-head belongs to a different cluster and also that a base-station is placed in the position of each cluster-head. So, the distance between a sensor and a base-station is measured in terms of number of hops. The objective function for cluster-head selection minimizes the number of base-stations through the minimization of the number of clusters. Thus, equation (4) is valid and subject to:

$$h(s_i, b_j) \leq h_{\max}, \forall\, s_i, b_j \in c_k \tag{9}$$

$$\forall\, s_i \in c_k, \exists\, s_j \in c_k \mid E(s_i, s_j) < R, i \neq j \tag{10}$$

The process starts by computing the number of neighbors of each sensor through the maximum number of hops allowed $h_{\max}$. This information is stored and the sensor with greatest number of neighbors is selected as a cluster-head. Then, it is checked if the cluster-head can reach all sensors. If not, the process of computing the number of neighbors per sensor repeats for the remaining sensors that are not reached by the last selected cluster-head until all sensors are in the reach of a cluster-head at $h_{\max}$.

2.5.2. Cluster Formation

In the cluster formation phase, overlapping clusters are disjoint. If a sensor is within the reach of more than one cluster-head, the sensor is assigned to the cluster of the closest cluster-head ($h_{\min}$), measured in the number of hops. In case there are more than one cluster-head at the same hop-distance, the sensor is assigned to the cluster-head with lowest weight ($w_{\min}$), i.e. smallest number of sensors in the cluster.

2.5.3. Dual Cluster-head Selection

After the cluster formation, the dual coverage phase takes place. Another computation of the number of neighbors is required for the sensors in the cluster that are not cluster-heads. Again, the sensor with the greatest number of neighbors is selected as cluster-head. If the new cluster-head can reach all neighbors of the cluster in $h_{\max}$, the cluster requires just two cluster-heads. Otherwise, it may be necessary to have more than two cluster-heads in the cluster in order to provide dual coverage. Finally, the base-stations can be defined in the same position of the cluster-heads of each cluster. If there is just one sensor in the cluster, just one base-station is required. We have:

$$|B_{c_k}| = \begin{cases} 1 & \text{if } |S_{c_k}| = 1 \\ 2 \text{ or more} & \text{if } |S_{c_k}| > 1 \end{cases} \quad (11)$$

The pseudo-code for the Dual DIS is given in Table 5.

Table 5. Pseudo-code for Dual DIS

Input: Number of sensors *N*, sensor positions (x_i, y_i), maximum number of hops $h_{\max}$.

Phase I - Cluster-head Selection

1. Initialization: $k = 1, j = 1$.
2. Assign all s_i to c_k.
3. **For** each $s_i \epsilon\ c_k$
    4. Compute neighbors of s_i in the range of $h_{\max}$.
5. **End for**
6. Define cluster-head $u_j = s_i$ with maximum number of neighbors.
7. **If** $\exists\ s_i \mid h(s_i, u_j) > h_{\max}$
    8. Remove s_i from c_k.
    9. $k = k + 1, j = j + 1$.
    10. Assign s_i to c_k.
    11. Go to 3.
12. **End If**

(continued)

Table 5. Pseudo-code for Dual DIS (continued)

Phase II - Cluster Formation

1. **For** $i = 1$ to N
    2. **If** ∃ more than one $u_j \mid h(s_i, u_j) \leq h_{\max}$
        3. Discard s_i from c_k with $h > h_{\min}$and assign s_i to c_k with $h = h_{\min}$.
        4. **If** ∃ more than one $u_j \mid h(s_i, u_j) = h_{\min}$
            5. Discard s_i from c_k with $w > w_{\min}$ and assign s_i to c_k with $w = w_{\min}$.
        6. **End if**
    7. **End if**
8. **End for**

(continued)

Table 5. Pseudo-code for Dual DIS (continued)

Phase III - Dual Cluster-head Selection

1. **For** $k = 1$ to M
    2. **For** $s_i \epsilon\ c_k$ and $s_i \neq\ u_j$
        3. Compute neighbors of s_i in the range of $h_{\max}$.
    4. **End for**
    5. $j = j + 1$.
    6. Define cluster-head $u_j = s_i$ with maximum number of neighbors.
    7. **If** $\exists\ s_i \mid h(s_i, u_j) > h_{\max}$
        8. Consider only $s_i \mid h(s_i, u_j) > h_{\max}$.
        9. Go to 3.
    10. **End if**
    11. **For** each $u_j \epsilon\ c_k$
        12. Define $b_j = u_j$, store (x_j, y_j).
    13. **End for**
14. **End for**

Output: Base-stations b_j, base-station placement locations (x_j, y_j).

CHAPTER 3. SIMULATION ANALYSIS

In this chapter, we show some simulation examples and we compare the performance of the algorithms. We use randomly placed sensors in a 10 x 10 square area and we run the simulations in MATLAB.

3.1. Examples

In this section, the same random distribution of sensors is used for comparison purposes in the Centroid with Coarse-Grid Zoom algorithm and the Dual DIS with $h_{\max} = 1$, $h_{\max} = 3$ and $h_{\max} = 5$. Figure 12 shows this random distribution with 80 sensors, where the position of each sensor is labeled with its ID number.

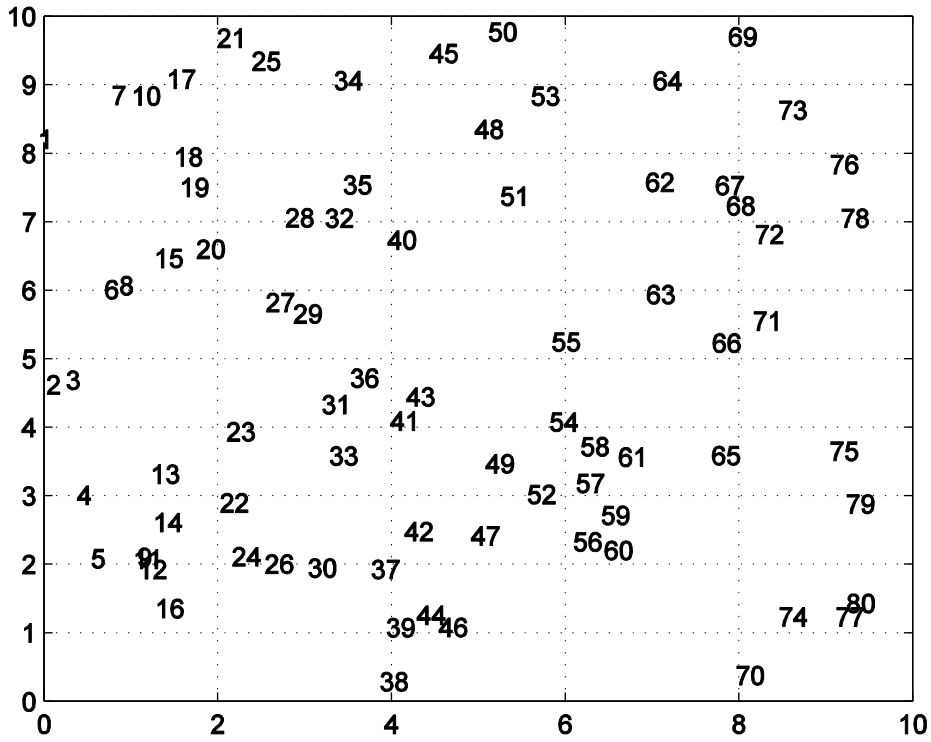

Figure 12. Random distribution of sensors

Figure 13 shows the Centroid with Coarse-Grid and Zoom algorithm where the random distribution of sensors from Figure 12. Random distribution of sensors

is divided in 25 clusters and 44 base-stations. The clusters are shown as dashed circles, the position of the sensors is labeled with the cluster number, and the position of the base-stations is indicated by squares. The base-station placement area is shown as the patch area inside each cluster. The sensors are connected to the closest base-station in the cluster by solid lines. There are 6 clusters with just one base-station (clusters with just one sensor) and 19 clusters with two base-stations.

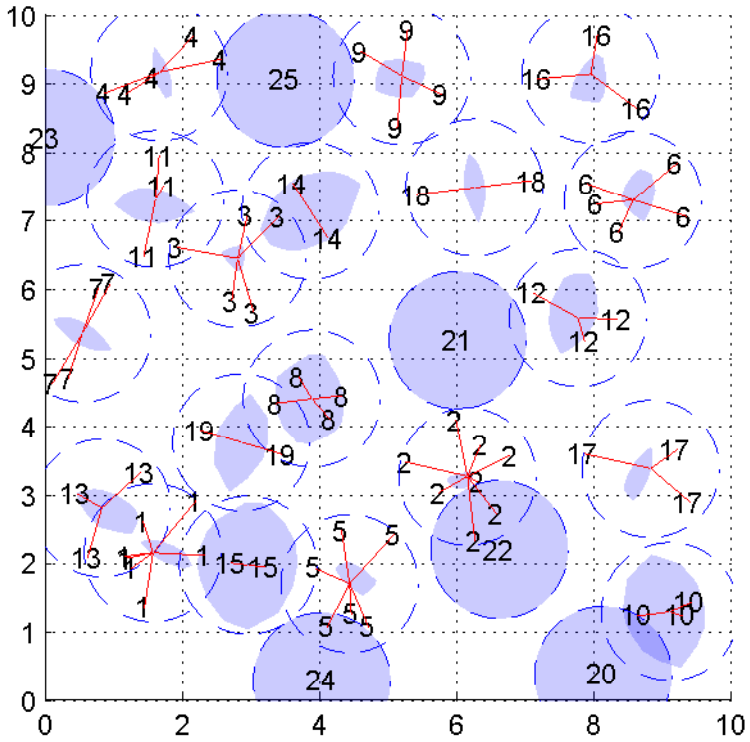

Figure 13. Centroid with Coarse-Grid and Zoom

Figure 14 shows DIS algorithm for $h_{\max} = 1$, resulting in 33 clusters, and 66 base-stations. A sensor that has not a base-station placed in the same position is identified by its ID number. The first base-station defined in each cluster is identified by the number of the cluster. Additional base-stations are identified by diamonds. For visualization, the elements of the same cluster are connected to the first selected base-

station using solid lines. Note that, in practice, each sensor connects to the closest base-station, or if load balance or other constraints are required, each sensor connects to the base-station with minimum constraints, in an ad-hoc configuration. Clusters with just one sensor, just have one base-station. Clusters with two sensors have two base-stations. Clusters with more than two sensors may have more than two base-stations.

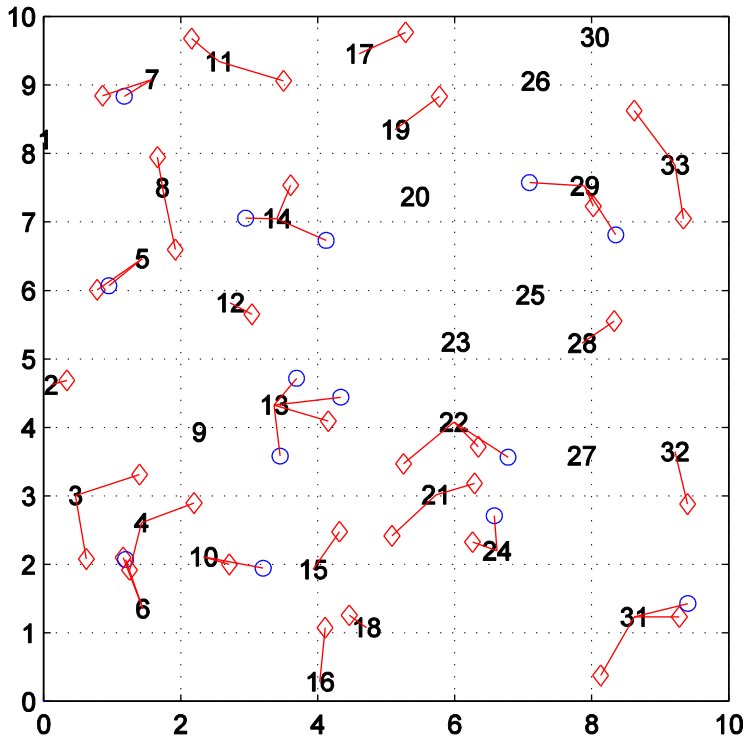

Figure 14. DIS for h_max=1

DIS is greed. In Figure 14, we can notice cluster 16 and 18 very close to each other, which indicate that they could be part of the same cluster. Using the labels of Figure 12. Random distribution of sensors

, sensor 39 could be the cluster-head of sensors 37, 38, 39, 44, 46, but following the criteria of number of neighbors, sensor 37 is selected as cluster-head of

sensors 30, 37, 39, 42, and 44 in the 4th iteration, sensor 46 is selected as cluster-head of sensors 39, 44, 46 in the 16th iteration, and sensor 38 is selected as cluster-head of sensors 38, and 39 in the 21st iteration. As the next criteria create disjoint clusters by associating the points that are at same hop-distance to the cluster-head with smaller weight, sensor 39 is put together to sensor 38 in cluster 16, and sensor 44 is assigned to cluster 18.

Figure 15 shows DIS algorithm for $h_{\max} = 3$ (24 clusters and 40 base-stations). The same conventions of the last figure are adopted.

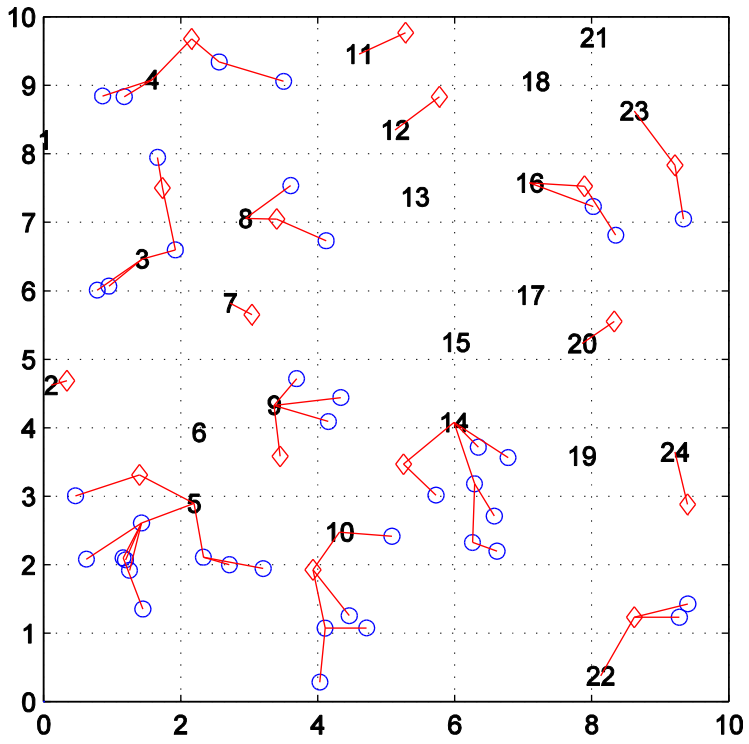

Figure 15. DIS for h_max=3

Figure 16 shows DIS algorithm for $h_{\max} = 5$ (23 clusters and 38 base-stations). The same conventions of the last figure are adopted.

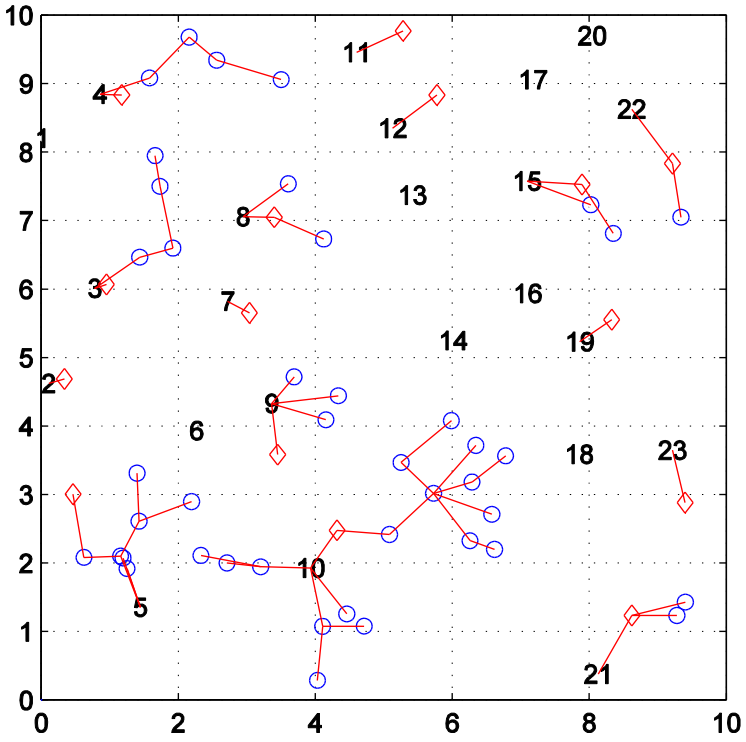

Figure 16. DIS for h_max=5

Comparing the Centroid algorithm and the DIS algorithm for this example, if we use $h_{\max} = 1$, the Centroid method has better performance than DIS. DIS gives 1.32 and 1.57 times more clusters and base-stations respectively than Centroid. Therefore, as the number of hops increase ($h_{\max} = 3$ and $h_{\max} = 5$), Centroid has a performance inferior than the DIS algorithm. Nevertheless, Centroid algorithm has the advantage that as any sensor can communicate with both of the base-stations, a failure in a sensor does not impact in the communication between the other sensors and the base-stations, while in DIS algorithm, for hops greater than one, this may not be valid. Also, as all sensors are able to communicate to the base-stations in each cluster, it is possible to configure a cluster-head rotation in Centroid and in DIS for $h_{\max} = 1$, while the cluster-head rotation cannot be applied in DIS for $h_{\max} > 1$. As one of our assumptions is that each sensor can communicate directly with at

least two base-stations, we use the Centroid and the DIS with $h_{\max} = 1$ in the massive simulations for comparison purposes in section 3.2.

3.2. Results

To measure the performance of the algorithms, we run 100 simulations for each clustering algorithm. We use the Centroid with Coarse-Grid and Zoom, and the Dual DIS with 1-hop algorithms for 50, 100, 200 and 300 randomly placed sensors cases. The metrics analyzed are the number of clusters, the number of base-stations, the average transmission distance between sensor and base-station, and the average transmission power between sensor and base-station.

Figure 17 shows the results for the number of clusters metric for R = 1. Besides the Centroid and DIS with one-hop clustering algorithms, we have the Grid Installation. method to provide a comparison for the number of clusters In Figure 17, the Centroid clustering algorithm provides consistently better results than DIS, giving from 33% to 18% less clusters for 50 to 300 sensor distribution cases respectively. This is largely due to the fact that in the Centroid method, there is no restriction for the base-station placement, while in DIS base-stations can only be placed in cluster-head positions. The average number of clusters for the 300 hundred sensors case using the DIS algorithm is above the Grid Installation, which is the result of the cluster overlaps, and shows the trend in the saturation of the number of clusters for higher density sensor distribution. In this situation, the direct application of the Grid Installation is more advantageous.

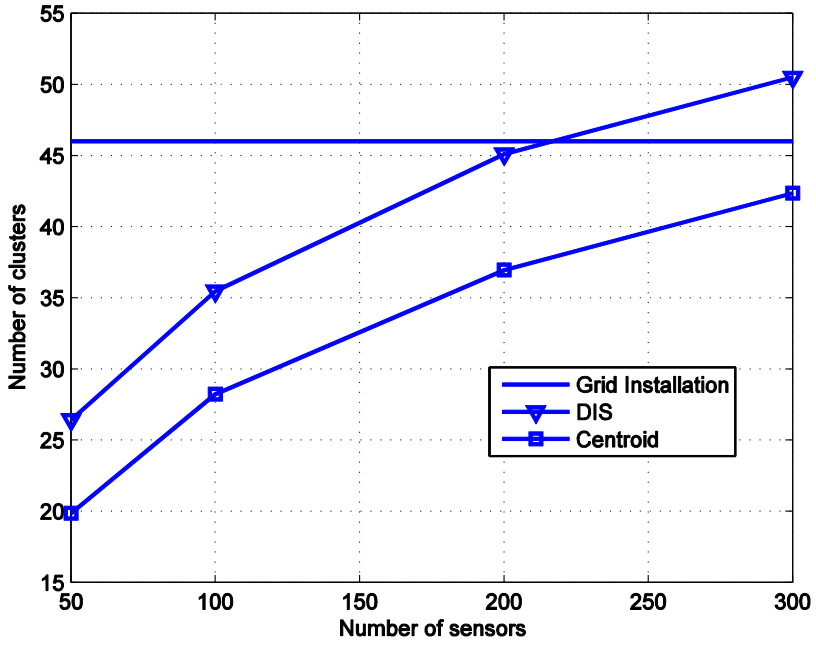

Figure 17. Number of clusters per clustering algorithm

Figure 18 and Figure 19 show the behavior of the algorithms in terms of number of clusters and number of base-stations metrics respectively for different radius length.

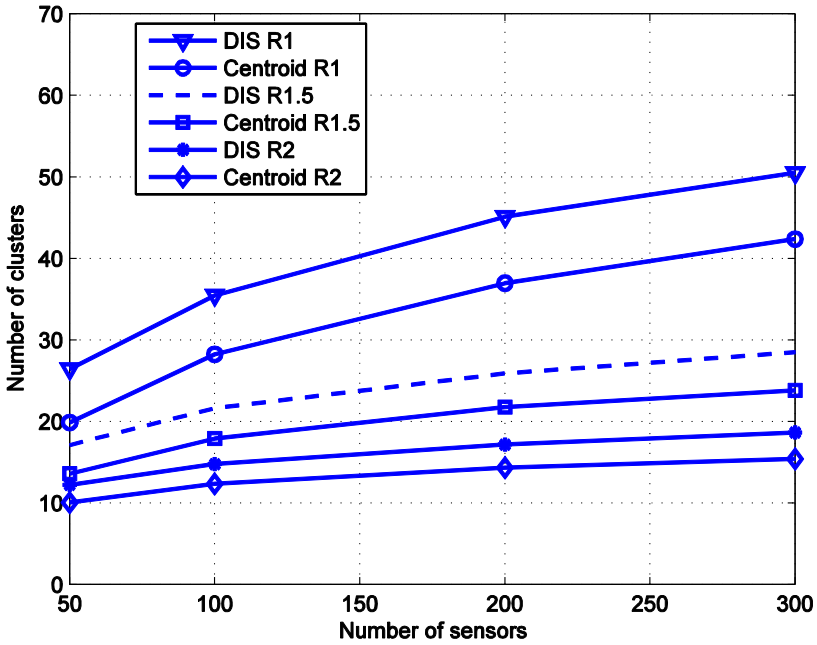

Figure 18. Number of clusters metric for different radius length

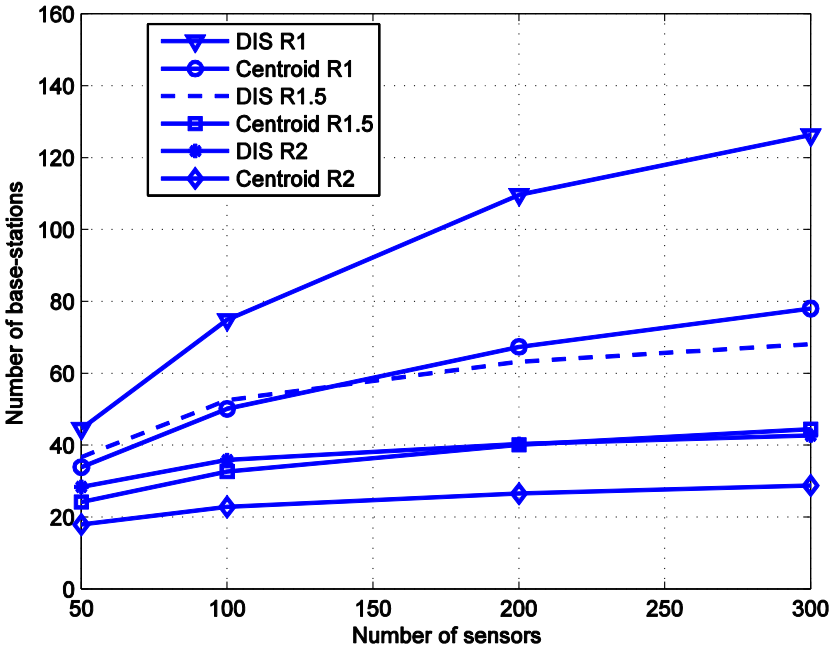

Figure 19. Number of base-stations metric for different radius length

As expected, the number of cluster and base-stations decrease as the radius increase regardless of algorithm. Also, Centroid algorithm always presents better performance than DIS for these metrics. While the curves of Centroid algorithm are shifted up by two units from number of cluster to number of base-stations metric, we do not observe the same proportion in the curves of DIS. Actually, the difference in proportion increases substantially as the coverage radius decreases. This is because when using DIS, we may need more than 2 base-stations per cluster to provide the dual coverage and an addition of a few more base-stations has a higher impact when smaller number of sensors is considered.

Figure 20 and Figure 21 show the results for the average transmission distance and average transmission power metrics. There are two scenarios per coverage radius length:

i. Coarse-Grid and Zoom: the two base-stations are placed inside the placement area of the cluster, but in different grid points, providing geographic diversity.

ii. Dual DIS with one-hop: the base-stations are placed in the same position of two different selected cluster-heads, providing geographic diversity.

For the transmission power metric we are considering $P_t = r^2$, where r is the average distance between sensor and base-station.

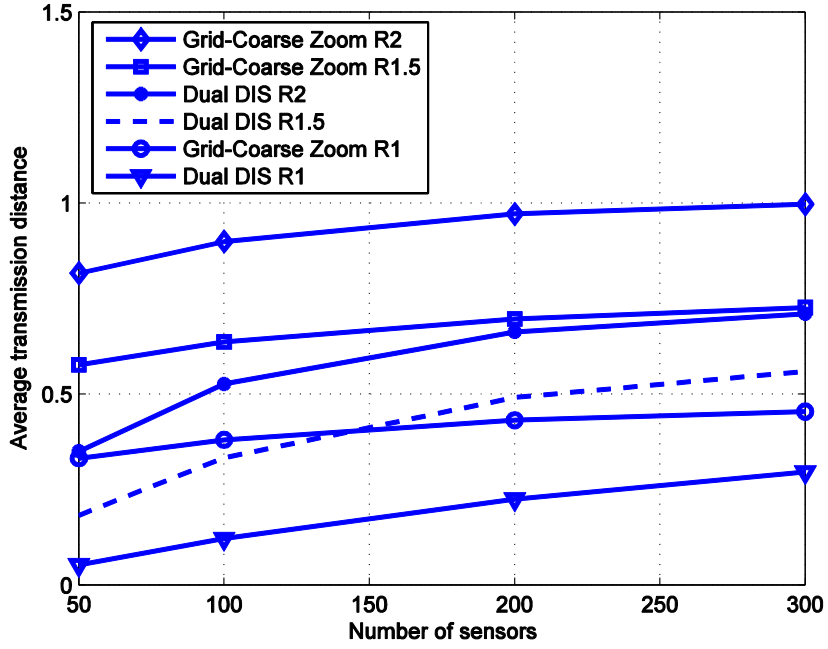

Figure 20. Average transmission distance metric for different radius length

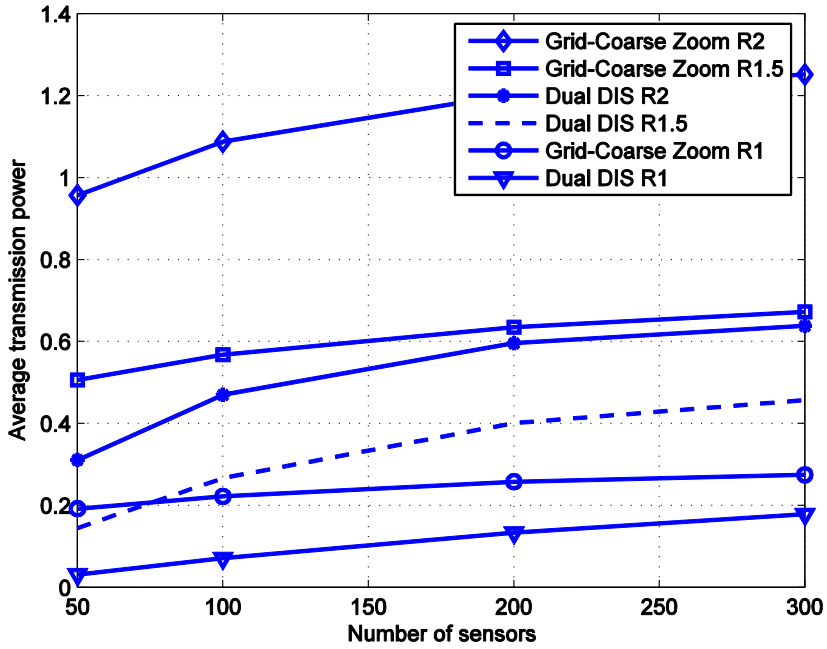

Figure 21. Average transmission power metric for different radius length

Dual DIS for $h_{\max} = 1$ provides lower average transmission distances and average transmission power when compared to Centroid with Coarse-Grid and Zoom method. This is the result of the trade-off in terms of number of base-stations when compared to the Centroid with Coarse-Grid and Zoom. This relies on the fact that in DIS there are always sensors directly connected to base-stations in every cluster, the ones that work as cluster-heads, as the base-stations are placed in the same position of the cluster-heads.

The behavior in Figure 19 is somewhat reverse than the behavior in Figure 20 and Figure 21. Centroid with Coarse-Grid and Zoom for R = 1 provides greater average transmission distances and average transmission power metrics than Dual DIS for R = 1.5 unit length (below 150 and 75 sensors respectively). As expected, the additional base-stations in the Dual DIS impacts positively in the minimization of transmitted power and negatively in the optimization of the total network cost.

In terms of time complexity, both algorithms are $\mathcal{O}(N^3)$ (see [6] for time-complexity of DIS). Centroid with Coarse-Grid and Zoom provides better number of clusters and number of base-stations metrics, while Dual DIS provides better average transmission distance and average transmission power metrics. Therefore, Centroid with Coarse-Grid and Zoom is the best method for the optimization of the cost of the network. If the solution requires a combination between the optimization of the cost of the network and the minimization of the power transmitted, the choice of the method depends on weight given on each of these factors.

CHAPTER 4. CONCLUSIONS AND FUTURE DIRECTIONS

This work presents two efficient heuristics to provide near-optimal solutions for dual coverage in wireless sensor networks: the Centroid with Coarse-Grid and Zoom, and the Dual DIS algorithms for clustering and base-station placement. It compares the results of the algorithms in terms of number of clusters, number of base-stations, transmission distance, and transmission power metrics. These are relevant metrics for this problem as the first two are related to the optimization of the total network cost and the others are related to the minimization of power consumption for battery life extension of the sensors.

The results show that there is a trade-offs between number of base-stations (clusters) and transmission distance/transmission power metrics. The Centroid with Coarse-Grid and Zoom method gives better results for the number of clusters/base-stations, while the Dual DIS method has better performance for the transmission distance/transmission power metrics.

In this work, we assumed that each sensor should be able to communicate directly with at least two base-stations, which implies that, for the clustering phase, just a single-hop is allowed. This assumption can be revisited and made more flexible to allow multi-hop and be objective of further investigation.

Future work can add other constraints besides the constraints of maximum number of hops allowed, such as, weight (maximum number of sensor per cluster) and relay load (maximum number of relay per sensor) constraints. The criteria for the selection of base-station placement can be broadened to consider variable transmission distance, throughput, bandwidth, and error rate values per sensor.

Also, in this work a static and centralized model was developed. Static because the position of sensors are fixed, and centralized because the decision of the computation of the

number of base-stations and the determination of base-station placement need to be done by a central unit. Further research is required to extend the analysis to mobile wireless sensor networks to allow decentralized and dynamic configurations.

REFERENCES

1. B. Everitt, S. Landau, M. Leese, D. Stahl, “Cluster Analysis”, 5th edition.

2. B. Lin, P. Ho, L. Xie1, and X. Shen, “Relay Station Placement in IEEE 802.16j - Dual-Relay MMR Networks”, 2008 IEEE.

3. B. Lin, P. Ho, L. Xie, and X. Shen, “Optimal Relay Station Placement in IEEE 802.16j Networks”, August 12–16, 2007.

4. M. Unbehaun, BMW Group, M. Kamenetsky, Stanford University, “On the Deployment of Picocellular Wireless Infrastructure”, Dec 2003.

5. A. Chehri, and H. Mouftah, “An Efficient Cluster-head Placement for Hybrid Sensor Networks”, ADHOC-NOW, 2010.

6. Y. Bejerano, “Efficient Integration of Multi-hop Wireless and Wired Networks with QoS Constraints”, Mobicom 2002.

7. J. Suomela, “Computational Complexity of Relay Placement in Sensor Networks”, SOFSEM, 2006.

8. X. Cheng, D. Du, L. Wang, B. Xu, “Relay sensor placement in wireless sensor networks”, Wireless Netw, 2008.

9. B. Aoun, R. Boutaba, Y. Iraqi, G. Kenward, “Gateway Placement Optimization in Wireless Mesh Networks with QoS Constraits”, 2005.

10. G. Takara, K. Xu, H. Hassanein, “Efficient Coverage Planning for Grid-Based Wireless Sensor Networks”, IEEE,2007.

11. G. Lin, G. Xue, “Steiner Tree Problem with Minimum Number of Steiner Points and Bounded Edge-Length”, Information Processing Letters, 1999.

12. W. Choi, P. Shah, S. Das, “A Framework for Energy-Saving Data Gathering Using Two-Phase Clustering in Wireless Sensor Networks”, IEEE, 2004.

13. M. Chatterjee, S. Das, D. Turgut, “WCA: A Weighted Clustering Algorithm for Mobile Ad Hoc Networks”, Cluster Computing, 2002.

14. E. Lloyd and R. Liu, “Approximating the Minimum Number of Maximum Power Users in Ad hoc Networks”, ADHOC-NOW, 2006.

15. R. Neapolitan, K. Naimipour, “Foundations of Algorithms”, 4º Edition.

16. C. Balanis, “Antenna Theory – Analysis and Design”, 2 Edition.