

LP WAN Gateway Location Selection Using Modified K-Dominating Set Algorithm [★]

Artur Frankiewicz², Adam Glos¹, Krzysztof Grochla¹, Zbigniew Łaskarzewski²,
Jarosław Miszczak¹, Konrad Połys¹, Przemysław Sadowski¹ and Anna Strzoda¹

¹ Institute of Theoretical and Applied Informatics, Polish Academy of Sciences
Bałtycka 5, Gliwice, Poland,

aglos@iitis.pl, kgrochla@iitis.pl, jmiszczak@iitis.pl, kpolys@iitis.pl,
psadowski@iitis.pl, astrzoda@iitis.pl,

² AIUT Sp. z o.o.,
Wyczółkowskiego 113, Gliwice, Poland
afrankiewicz@aiut.com, zlaskarzewski@aiut.com

Abstract. The LP WAN networks use gateways or base stations to communicate with devices distributed on large distances, up to tens of kilometres. The selection of optimal gateway locations in wireless networks should allow providing the complete coverage for a given set of nodes, taking into account the limitations, such as the number of nodes served per access point or required redundancy. In this paper, we describe the problem of selecting the base stations in a network using the concept of k -dominating set. In our model, we include information about the required redundancy and spectral efficiency. We consider the additional requirements on the resulting connections and provide the greedy algorithm for solving the problem. The algorithm is evaluated in randomly generated network topologies and using the coordinates of sample real smart metering networks.

Keywords: LP WAN, radio planing, gateway location, redundancy

1 Introduction

The introduction of Low Power WAN network concept, with devices communicating over distances of tens of kilometers using low power radio interface, has generated multitude of novel use cases based on long life battery powered devices. The LoRa (Long Range) is one of the most widely adopted LP WAN standards, which is based on ISM (Industrial, Scientific and Medical) frequencies and chirp spectrum modulation. The LoRa communication now is being deployed in devices such as smart meters, sensors or actuators executing smart city functions. The LoRa can provide cheap and very energy efficient communication with thousands of devices per one access point, at the cost of low data rate. To facilitate the deployment of LoRa devices, the LoRaWAN standard has

[★] This work has been partially funded by the Polish National Center for Research and Development grant no. POIR.04.01.04-00-0005/17.

been proposed, which defines the packet format and communication architecture allowing to forward the information from the devices to the internet servers.

Traditionally radio planning technologies are used to select location of base stations in long range wireless networks. This approach is not always feasible in LP WAN, as due to the low cost of access point (few hundred dollars) and the simplicity of its deployment it is more efficient to deploy the access points in less efficient manner, but without the costly radio planning procedures. Traditionally, the gateway location selection problem was solved by radio planning engineers, using radio signal propagation models and simulation software and tested through the drive tests [1], [2]. This method is too expensive for the LP WAN networks, which employ very large number of low cost devices, as it requires both engineering time to plan the radio and execution of time consuming drive tests. The number and locations of gateways need to be dynamically modified basing on the growth of the network, to provide coverage appropriate to the endpoints' density, as more end nodes generate more traffic which needs to be served.

In this work, we propose an algorithm to select the suboptimal location for the LP WAN gateway location for a given set of endpoint location. The gateway location is selected from a set of location candidates, which may be the same as the set of end node location. The selection of a set of gateways must meet the requirement to provide connectivity to all end nodes and to provide the required redundancy. We assume that the connectivity graph estimated from the distances between the nodes is the input. We treat the problem as a problem of selection of a dominating set, which is an NP-complete decision problem in computational complexity theory [3]. Thus the problem is complex and it is hard to solve.

There are a few differences between other popular radio networks (e.g. LTE) and LoRa. LoRa end devices are typically battery powered devices so they transmit data rarely, but usually in a constant time window. LoRa uses ALOHA media access control so collisions are more probable. LoRa devices also use all possible channels uniformly, there is no "colouring" of areas and LoRa uses spreading factors, thus the capacity of the gateway may vary depending on the spatial distribution of the nodes. We assume the set of client locations and the traffic characteristics are known and is an input for the proposed algorithm, which is a case e.g. for advanced metering infrastructure (AMI) networks.

The rest of the paper is organized as follows: the second section describes the state of the art regarding the planning and dimensioning of LP WAN networks and the selection of AP location. Next, we describe the network model used in the study. In the following section, we show the pseudocode of the proposed optimization algorithm. Next we present the analysis of the performance of the algorithm for different types of topologies. We finish the paper with a short conclusion.

2 Literature review

The radio network planning has been investigated for more than twenty years, however most of the work is dedicated to cellular networks - see e.g. [1] or [2]. However, there is very little research results available considering the specifics of the LP WAN radio network planning, where the cost of network deployment is much smaller, thus it does not justify the costly drive tests and radio signal propagation analysis. One of the few tools which allow to estimate the coverage of a LoRa gateway on a map is CloudRF [4]. There are some other tools allowing to simulate radio signal propagation, of which most support ISM frequencies and allow to estimate the range of LP WAN access points, such as e.g. Radio Mobile Online [5]. These tools are very useful in analysis of network coverage for a given set of access points, but require large knowledge of RF signal propagation and do not select automatically the gateway location, requiring the experience of the operator. While this is effective for a small number of access points, for large networks which need to match the restraints of number of nodes services by an access point use of such tools is ineffective. In the previous work by part of the authors [6] two of us have proposed a heuristic algorithm for gateway location selection, which however do not support redundancy.

3 Gateway location selection procedure

3.1 Problem formulation

To describe the problem of selecting the set of gateways for the given set of stations (end points), we assume that we have for our disposal the graph describing the visibility between the stations. In such graph two stations are connected only if it is possible to transmit packets between them. We assume that different connections may have different costs. We also make some assumption on capabilities of the resulting network. First, we assume that each gateway has a fixed maximal capacity, the same for each of them, representing the maximal number of stations it can connect to simultaneously. This ensures that the transmission load will not exceed the capacity of the gateway. Second, we assume that each non-station has to be connected at least to some fixed number of gateways. This condition describes the redundancy requirements for the resulting network. Finally, we would like to find the smallest possible set of gateway, which expresses the requirement of using the smallest possible number of gateways.

To represent this situation we use a triple (G, c, k) , where G is undirected weighted graph $G = (V, E)$, with the set of vertices V and the set of edges E , c is the maximal capacity of the gateways, and k is the domination number describing the minimal number of gateways each station is connected to. In this description each vertex represents a station, and stations are connected if they are within their radius of visibility. To each edge, $(v, w) \in E$, we assign a weight C_{vw} , representing the cost of the connection between the station represented by vertices v and w . The connection cost between nodes v and w is equal to

$\frac{1}{2^{12-SF_{vw}}}$, where SF_{vw} is equal to the value of spreading factor in the communication between nodes v and w . Values of SF in the communication between two nodes are determined on the basis of a signal propagation model that takes into account the distance between nodes and random attenuation. The higher the SF value, the higher the transmission cost. The domination number k is used for representing the required redundancy.

3.2 Description of the solution

To describe our approach we first consider a simplified sub-problem, defined as follows.

Problem 1 (Gateways). For a given set of stations, described by (G, c, k) , find the set D of stations such that each station is in this set or it has a direct connection to a station from this set.

One can note that, as the only requirement here is to assure the visibility between the stations, the solution of the above problem is given a subset of stations. Thus, no information about the connections is given.

The stations from the set D , obtained as a solution of the above problem, are called gateways. The set D is called a dominating set of G . If we impose the condition that each element station should be connected to at least k gateways, the solution of the above problem is equivalent to finding k -dominating set, D_k , of a graph G .

Definition 1. For a graph $G = (V, E)$, the k -dominating set of G , denoted by D_k , is any subset of vertices such that each vertex from V is connected to at least k nodes from D_k or belongs to D_k .

The notion of k -dominating set captures the redundancy requirement. Still the information about the D_k alone is not sufficient to describe the allocation of connection between stations and gateways.

A similar approach has been used previously for constructing routing algorithms in *ad hoc* networks [7], and various methods have been developed in this context [8–10]. However, one should keep in mind that in our scenario we assume that not only the resulting network has to be redundant, but also the resulting gateways have limited capacities. On the other hand, in contrast to the routing problem, we do not require the subgraph with vertex set consisting of gateways to be connected.

One should note, that if we assume the capacity c in our problem to be infinite, then an arbitrary connection scheme with a gateway D_k would be a proper solution. In particular, gateway can serve all of the neighbouring nodes. However, since capacity is finite, each gateway cannot serve all neighbouring nodes, thus we may need to add gateways in order to serve not-served one. Thus, the solution of our problem is given by the solution of k -domination problem, which is to find the minimal k -dominating set, with fixed capacity c .

Taking all this into account, we define our problem as follows.

Algorithm 1 Greedy algorithm for finding k -dominating scheme and its connection scheme. The stop condition is that every vertex is either a gateway or is served by k gateways. If the stop condition is not fulfilled, new gateway according to NEW_GATEWAY is chosen. Here G is input graph, C is the cost function, k is the domination parameter and c is the common capacity for all gateway nodes

```

1: function CREATE_CONNECTION_GRAPH( $G = (V, E), C, k, c$ )
2:    $E_H \leftarrow \emptyset$ 
3:    $D \leftarrow \emptyset$ 
4:    $H := (V, E_H, f)$ 
5:   while  $\{v \in V : \deg_H(v) < k\} \setminus D \neq \emptyset$  do
6:      $w, S \leftarrow \text{NEW\_GATEWAY}(G, H, C, D, k, c)$ 
7:      $D \leftarrow D \cup \{w\}$ .
8:     for all  $v \in S$  do
9:        $E_H \leftarrow E_H \cup \{\{v, w\}\}$ 
10:  return  $D, H$ 

```

Problem 2 (Gateways with constraints). For a given graph, $G = (V, E)$, find the set of gateways D_k , forming a k -dominating set $D_k \subset V$, and a graph, $H = (V, E_H)$, which is a subgraph of G , $H \subset G$, satisfying the following conditions,

1. each edge from E_H connects gateways and stations,
2. each end point is connected to at least k gateways,
3. the sum of weights for edges for each gateway v is at most c ,

$$\sum_{w \in V, \{w, v\} \in E_H} C_{vw} \leq c. \quad (1)$$

The last condition assures that the redundancy requirement is fulfilled, and even if for given station $k - 1$ neighbouring gateway would be unavailable, it is promised that k th gateway will have enough resources for serving the station.

3.3 Greedy algorithm

Let $G = (V, E)$ be an undirected graph with $n = |V|$ vertices. The algorithm is an extension of the greedy algorithm for k -dominating set creation [11]. It is a greedy algorithm that starts with $D = \emptyset$, and adds a new vertex w minimizing current value of $a(G, D \cup \{w\})$ of the form

$$a(G, D) = nk - \sum_{v \in V} d_k(v, D), \quad (2)$$

where

$$d_k(v, D) = \begin{cases} \min(k, N(v) \cap D), & v \notin D, \\ k, & v \in D, \end{cases} \quad (3)$$

Algorithm 2 The algorithm which chooses best end point to be new gateway. It is chosen to be the end point, which can serve maximal possible stations. Here G is input graph, H is the subgraph of G , C is the cost function, D is the collection of gateways, k is the domination parameter and c is the common capacity for all gateway nodes. A degree of vertex v in graph H is denoted as $\deg_H(v)$.

```

1: function NEW_GATEWAY( $G = (V, E), H = (V, E_H), C, D, k, c$ )
2:   for  $w \in V$  do
3:     if  $w \in D$  then
4:        $\text{value}[w] \leftarrow 0$ 
5:     else
6:        $T \leftarrow N_G(w)$   $\triangleright N_G(w)$  returns neighbouring nodes
7:        $T \leftarrow T \setminus D$   $\triangleright$  remove gateways
8:        $T \leftarrow \{v \in T : \deg_H(v) < k\}$   $\triangleright$  remove served nodes
9:        $S \leftarrow \emptyset$ 
10:      while  $\sum_{v \in S} C_{vw} \leq c$  do  $\triangleright$  check if the station can still serve more
         stations
11:         $v \leftarrow \arg \min_{v \in T} C_{vw}$   $\triangleright$  choose neighbour with smallest connection
         cost
12:         $S \leftarrow S \cup \{v\}$ 
13:         $T \leftarrow T \setminus \{v\}$ 
14:         $\text{value}[w] \leftarrow 1 + |S|$   $\triangleright$  because marking  $w$  makes it served
15:       $w \leftarrow \arg \max_{v \in V} \text{value}[v]$ 
16:    return  $w, S$ 

```

where $N(v)$ is the neighbourhood of v in G . The algorithm stops when $a(G, D) = 0$, which implies that D is a k -dominating set. Note that function $a(G, D)$ describes how many nodes are already in D , or in the case of non-gateways how many gateways are within their radius.

In our scenario the gateways, which are elements of a k -dominating set, have limited capacity. Thus, we need to store which end points connect to newly added gateway. Hence, we start with empty gateway collection D and empty graph $H = (V, E_H = \emptyset)$.

We choose gateways until (D, H) is a proper solution of our problem. The procedure to provide the solution is provided in Algorithm 1. At each step, new gateway is chosen to be the one maximizing the number of end point stations that it can serve and that requires service S . The details are provided in Algorithm 2. When the gateway w is chosen, it is added to current gateway collection $D = D \cup \{w\}$, and the edges set E_H is updated with edges connecting w and nodes it promised to serve by $E_H = E_H \cup \{\{w, v\} : v \in S\}$.

4 Performance evaluation

The proposed algorithm has been evaluated using the randomly generated topologies and the real topologies of smart metering networks in two Polish cities which

datasets we have access to, but are not in public domain. The algorithms are compared primarily in terms of the number of base stations selected for a given network topology, method operation time and spreading factors in the communication between selected base stations and end devices assigned to them by each of two algorithms. Experiments have been carried out for algorithm parameter sample values such as $k \in \{1, 2, 3\}$ and fixed maximal capacity of the gateway. The value of the maximal capacity parameter has been determined on the basis of our calculations taking into account transmission parameters such as the probability of packet delivery, the transmission time of a single packet and the transmission window.

As a reference method, we've used an implementation of the algorithm from [12], as the most similar to our approach. To be more precise, we've applied the heuristic method named accelerated dynamic weighted greedy algorithm (ADWGA) which is algorithm edition based on the recommendations of the authors of [12] this version of the algorithm for large-scale cases. In this approach, the base stations are not limited to the maximum number of end devices that can be assigned to a single base station. The authors of [12] do not provide exact values of input algorithm parameters, marked in [12] as α and β , but only appoint the ranges within which the parameter values should be. Searching for the reference method parameter values has a significant impact on the algorithm's operation time. Due to the long-running calculations of the reference method, we conducted the experiments using the reference method [12] for two real-life network topologies and for some small random data sets. The results of both methods are compared with each other in table 1 (for real network topologies) and table 2 (for random network topologies).

Figure 1 depicts the distribution of nodes for two real-life IoT deployment topologies of two sample cities A and B, being the subject of experiments. City A is a typical small city topology and city B consists of nodes divided in two separate regions – a double city estate topology.

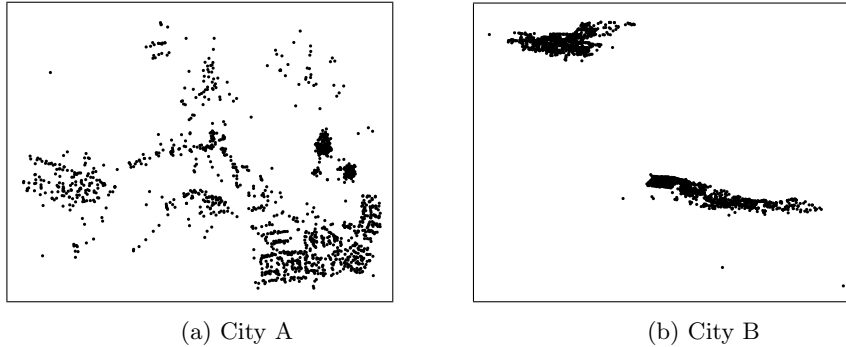


Fig. 1: Real-life IoT deployment network topologies of two sample cities A and B.

In table 1 the results of both greedy and reference method [12] for real network topologies are summarized. The results in the table 1 include an amount of selected gateways ('number of gateways'), algorithm operation time in which the result was obtained ('time [s]') and the average spreading factor to the gateway ('avg. SF'). Even though in most cases the reference algorithm found a smaller number of base stations than our method, the difference is very small. However, the difference in the time of operation of both methods is huge. For these two real network topologies our algorithm finds comparable number of base stations in much less time than reference method.

Table 1: Results of the proposed greedy algorithm and reference method [12] (ADWGA) for two real-life network topologies A and B.

| city | nodes | area [m] | k | number of gateways | | time [s] | | avg. SF | |
|------|-------|--------------------|---|--------------------|----------------|-------------|----------------|---------|----------------|
| | | | | greedy | reference [12] | greedy | reference [12] | greedy | reference [12] |
| A | 1719 | 2490×4090 | 1 | 6 | 6 | 4.95 | 253.08 | 9.19 | 9.18 |
| | | | 2 | 12 | 11 | 8.86 | 3118.27 | 9.2 | 8.92 |
| | | | 3 | 17 | 15 | 12.58 | 7603.15 | 9.17 | 8.88 |
| B | 1760 | 1260×3060 | 1 | 8 | 7 | 2.47 | 169.45 | 7.8 | 8.06 |
| | | | 2 | 15 | 14 | 4.57 | 363.16 | 7.84 | 7.86 |
| | | | 3 | 20 | 19 | 6.72 | 2249.87 | 7.84 | 7.89 |

Experimental results presented below include proposed algorithm performance statistics for various scenarios assuming the different number of network nodes randomly positioned in accordance with the uniform distribution in changing areas. The algorithm has been tested on network topologies with different density of nodes distribution in space. The statistics presented in tables below were obtained based on 30 runs of the algorithm for each number of nodes equals 1000, 2500, 5000, 10000 and 20000 distributed in a rectangle with dimensions $5000\text{m} \times 7500\text{m}$, $10000\text{m} \times 15000\text{m}$, $15000\text{m} \times 22500\text{m}$ and $20000\text{m} \times 30000\text{m}$. Before each running of the algorithm positions of nodes were selected randomly. Tables 3, 4 and 5 contain statistics such as the average number of selected gateways, the average algorithm running time and the average spreading factor occurring between gateways and their clients (end point nodes) for redundancy factor $k \in \{1, 2, 3\}$ respectively and with fixed maximal capacity.

In table 2 greedy and reference method statistics, described above, for small data sets are contained. It can be seen that in each case the running time of the greedy algorithm is much shorter than the running time of the reference method, especially for network topologies characterized by a dense distribution of nodes in space. In cases where the reference algorithm returns a smaller number of base stations than our method, which is on the plus side for the ADWGA, because it generates lower costs, then the greedy algorithm returns a little more base stations in a much shorter time. Moreover, for less dense network topologies our algorithm finds significantly fewer base stations much faster than reference

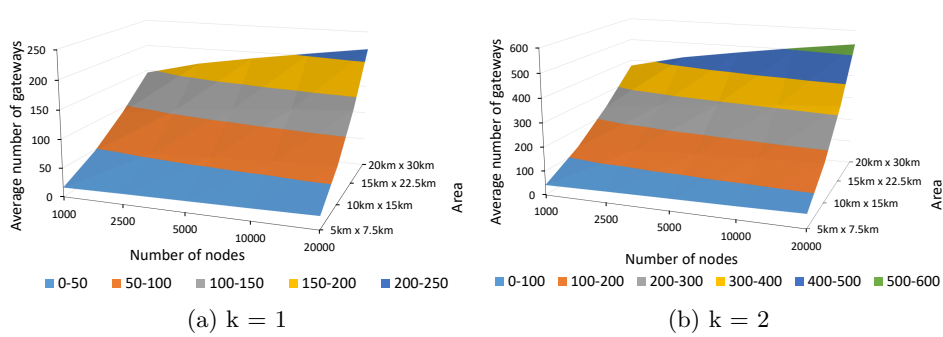
Table 2: Results of the greedy and reference [12] method for sample topology scenarios.

| nodes | area [m] | k | avg. number of gateways | | avg. time [s] | | avg. SF | |
|-------|---------------|---|----------------------------|----------------|------------------|----------------|---------|----------------|
| | | | greedy | reference [12] | greedy | reference [12] | greedy | reference [12] |
| 1000 | 5000 × 7500 | 1 | 16.31 | 15.37 | 0.75 | 71.74 | 9.52 | 9.44 |
| | | 2 | 29.9 | 27.5 | 1.41 | 640.51 | 9.54 | 9.58 |
| | | 3 | 42.38 | 38.31 | 2.03 | 1139.58 | 9.57 | 9.56 |
| 1000 | 10000 × 15000 | 1 | 47.66 | 47.94 | 0.23 | 64.58 | 9.62 | 9.61 |
| | | 2 | 86.03 | 83.2 | 0.43 | 733.19 | 9.67 | 9.66 |
| | | 3 | 123.34 | 117.69 | 0.61 | 933.72 | 9.7 | 9.68 |
| 1000 | 15000 × 22500 | 1 | 91.34 | 92.59 | 0.13 | 80.6 | 9.69 | 9.69 |
| | | 2 | 166.45 | 166.41 | 0.25 | 939.71 | 9.74 | 9.71 |
| | | 3 | 236.34 | 249.41 | 0.34 | 1074.58 | 9.77 | 9.7 |
| 1000 | 20000 × 30000 | 1 | 145.38 | 153.93 | 0.1 | 348.31 | 9.74 | 9.74 |
| | | 2 | 264.31 | 296.03 | 0.19 | 739.93 | 9.78 | 9.71 |
| | | 3 | 373.38 | 566.28 | 0.27 | 896.88 | 9.78 | 10.02 |
| 2500 | 5000 × 7500 | 1 | 18.14 | 16.59 | 5.4 | 624.57 | 9.46 | 9.38 |
| | | 2 | 32.52 | 30.24 | 9.94 | 8496.08 | 9.48 | 9.52 |
| | | 2 | 46.72 | 41.14 | 14.39 | 18170.28 | 9.5 | 9.5 |

method. The speed of the greedy algorithm allows us to obtain results for large-scale network topologies in a relatively short time.

Figure 2 is the graphic interpretation of results for the average number of selected gateways presented in table 2. To make the charts more transparent, the dimensions of the network topology area are presented in units of kilometers instead of meters, as in the tables.

On figures 3, 4 and 5 a graphic interpretation of the greedy method has been shown to illustrate the results presented in the tables 2, 3, 4, 5 for $k \in \{1, 2\}$.


 Fig. 3: Average number of selected gateways for each scenario for redundancy factor $k \in \{1, 2\}$.

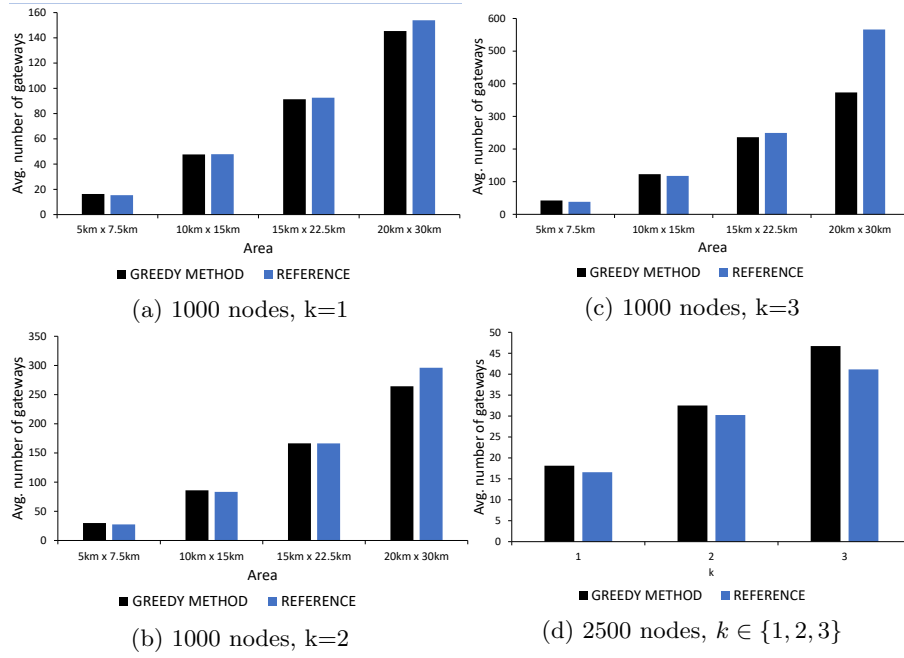


Fig. 2: Comparison of the average number of selected gateways for $k \in \{1, 2, 3\}$ for both methods; the proposed greedy algorithm and reference method [12] for 1000 number of nodes deployed in all areas (a), (b), (c) and for 2500 number of nodes deployed in the smallest of considered areas (d).

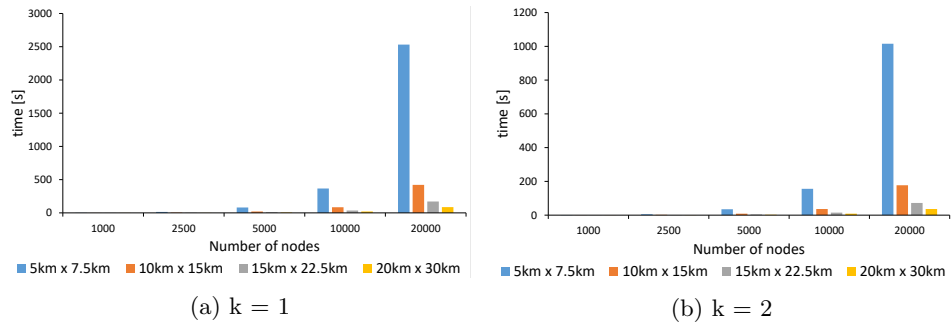


Fig. 4: Average greedy algorithm run time for each scenario for redundancy factor $k \in \{1, 2\}$.

Table 3: Statistics of the proposed greedy algorithm of finding k – dominating set for $k = 1$.

| nodes | area [m] | avg. number of gateways | avg. time [s] | avg. SF |
|-------|---------------|----------------------------|------------------|---------|
| 2500 | 10000 × 15000 | 53.9 | 1.97 | 9.54 |
| | 15000 × 22500 | 105.9 | 0.97 | 9.64 |
| | 20000 × 30000 | 170.6 | 0.61 | 9.7 |
| 5000 | 5000 × 7500 | 19.53 | 34.79 | 9.43 |
| | 10000 × 15000 | 58.73 | 8.74 | 9.49 |
| | 15000 × 22500 | 115.53 | 4.06 | 9.57 |
| | 20000 × 30000 | 188.03 | 2.51 | 9.64 |
| 10000 | 5000 × 7500 | 20.97 | 155.8 | 9.41 |
| | 10000 × 15000 | 62.53 | 36.39 | 9.45 |
| | 15000 × 22500 | 124.4 | 15.07 | 9.52 |
| | 20000 × 30000 | 203.8 | 8.78 | 9.58 |
| 20000 | 5000 × 7500 | 22.73 | 1015.3 | 9.21 |
| | 10000 × 15000 | 67.1 | 177.7 | 9.43 |
| | 15000 × 22500 | 132.7 | 72.3 | 9.47 |
| | 20000 × 30000 | 218.9 | 36.3 | 9.52 |

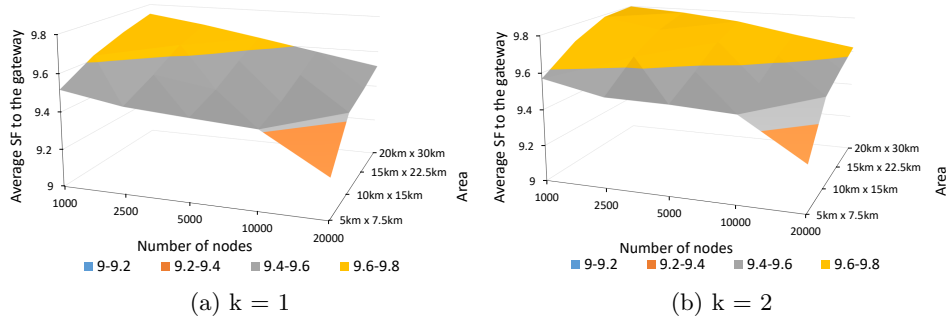


Fig. 5: Average Spreading Factor to the gateway for each scenario for redundancy factor $k \in \{1, 2\}$.

The distribution plots shown in the figure 6 present how the percentage number of non – gateway having a given spreading factor to the gateway is distributed for a sample consisting of specified network topologies. A single sample consists of 150 topologies with an amount of nodes equal to 1000, 2500, 5000, 10000, 20000 distributed in a given area, where 30 random topologies were generated for each of the five number of nodes. Each of the plots 6a, 6b, 6c, 6d was created on the basis of described 150 topologies generated for each area separately.

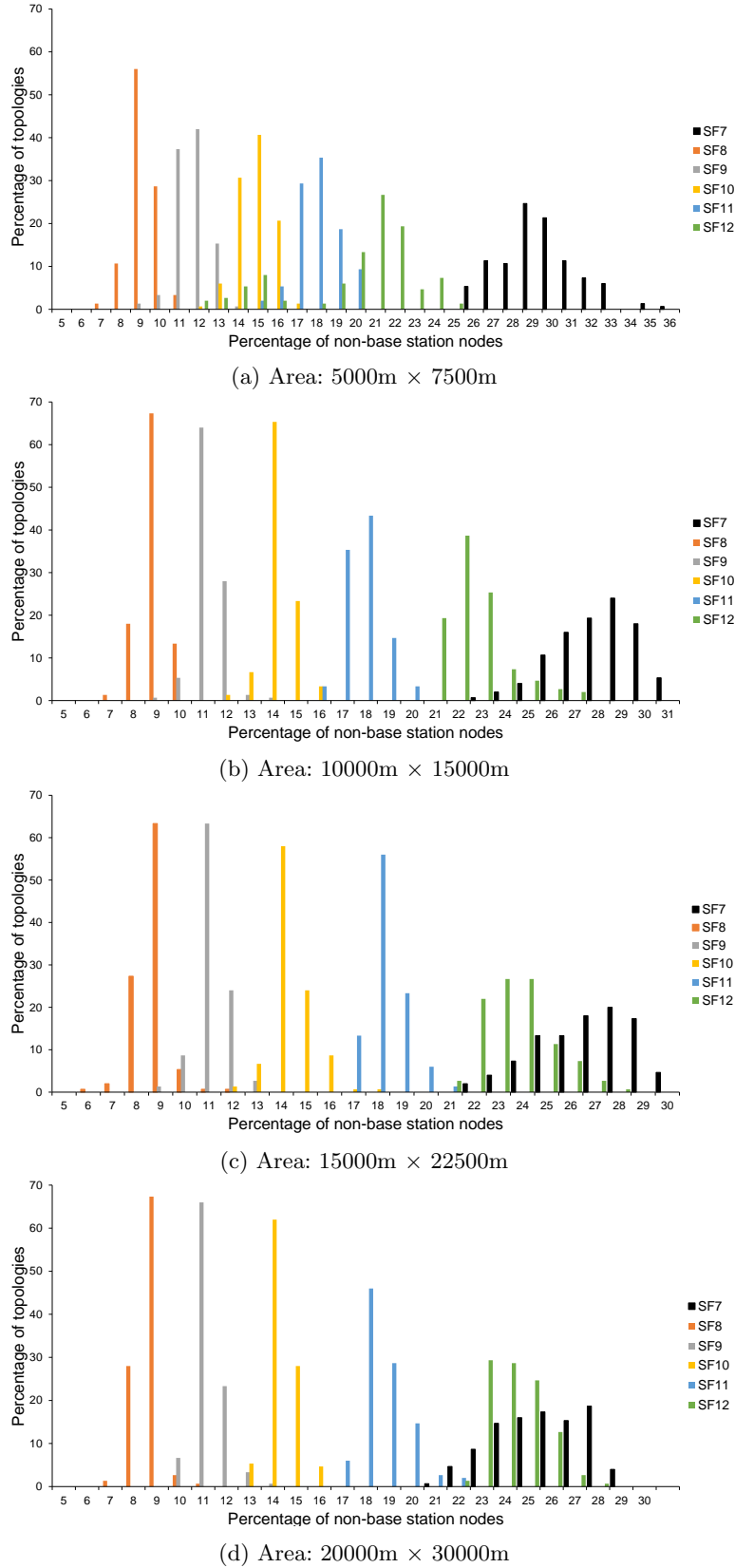


Fig. 6: Distributions of the percentage number of nodes having a specific SF to the gateway for different sizes of the node distribution area.

Table 4: Statistics of the proposed greedy algorithm of finding k – dominating set for $k = 2$.

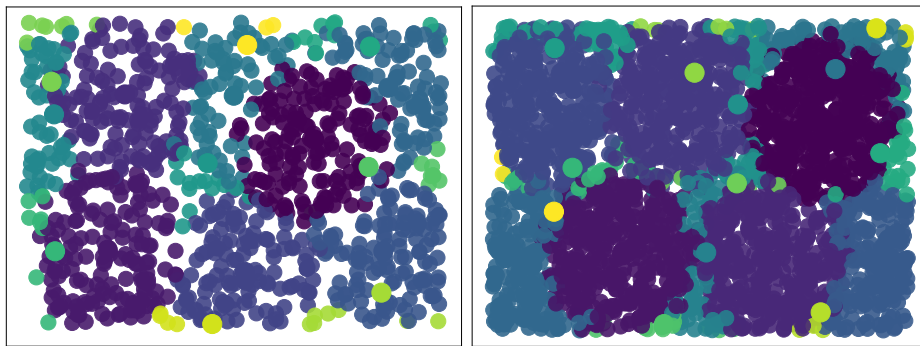
| nodes | area [m] | avg. number of gateways | avg. time [s] | avg. SF |
|-------|----------------------------|----------------------------|------------------|---------|
| 2500 | 10000×15000 | 97.1 | 3.41 | 9.61 |
| | 15000×22500 | 188.43 | 1.60 | 9.7 |
| | 20000×30000 | 305.07 | 1.08 | 9.75 |
| 5000 | 5000×7500 | 35.43 | 59.42 | 9.48 |
| | 10000×15000 | 104.80 | 14.84 | 9.55 |
| | 15000×22500 | 206.03 | 7.07 | 9.65 |
| | 20000×30000 | 333.27 | 4.31 | 9.71 |
| 10000 | $5000 \times 7500\text{m}$ | 37.57 | 264.52 | 9.45 |
| | 10000×15000 | 111.63 | 61.24 | 9.52 |
| | 15000×22500 | 220.3 | 25.55 | 9.59 |
| | 20000×30000 | 361.53 | 14.91 | 9.65 |
| 20000 | $5000 \times 7500\text{m}$ | 40.83 | 1735.58 | 9.25 |
| | 10000×15000 | 117.23 | 301.56 | 9.48 |
| | 15000×22500 | 233.81 | 121.46 | 9.54 |
| | 20000×30000 | 385.35 | 61.6 | 9.59 |

The distributions of percentage number of non – gateway nodes having a spreading factor equal to SF8, SF9, SF10 or SF11 to the target gateway are in very similar ranges of values, regardless of the size of the area. For instance, the percentage of the number of nodes having SF8 to gateway from topologies distributed on area $5000\text{m} \times 7500\text{m}$ is in the range from 6% to 11% and stays within similar limits despite the increasing area and, consequently, the decreasing network density. The same phenomena can be observed for percentage of number of nodes that have spreading factor to gateway equals to SF9, SF10 and SF11. The algorithm selects the locations of gateway in such a way that the percentage of clients with a given SF to the gateway is very similar, regardless of the density of the network. Around each gateway, there are areas within there are non – gateway nodes having a given spreading factor SF7, SF8, etc. The higher the density, the proportionally higher percentage of nodes is in the area with spreading factor SF7 and less in the area with SF12. Similarly, the lower the density, the proportionally smaller percentage of nodes falls into the area with spreading factor SF7 to the gateway and more into the area with spreading factor SF12. In the remaining areas, the number of nodes in each of them is proportionally comparable regardless of the density of the network. In the figure 6 can be observed that as the area increases and, as a consequence, the network density decreases, the values of the bins in distributions of nodes with SF7 to the gateway decrease. In the subsequent charts, the percentage range of nodes having SF7 to the gateway moves left along the horizontal axis, simultaneously the percentage range of nodes with SF12 to the gateway moves right along the same

Table 5: Statistics of the proposed greedy algorithm of finding k – dominating set for $k = 3$.

| nodes | area [m] | avg. number of gateways | avg. time [s] | avg. SF |
|-------|----------------------|----------------------------|------------------|---------|
| 2500 | 10000×15000 | 137.43 | 4.62 | 9.58 |
| | 15000×22500 | 266.5 | 2.21 | 9.68 |
| | 20000×30000 | 430.87 | 1.54 | 9.74 |
| 5000 | 5000×7500 | 50.27 | 81.38 | 9.49 |
| | 10000×15000 | 147.63 | 20.66 | 9.58 |
| | 15000×22500 | 289.57 | 9.82 | 9.68 |
| | 20000×30000 | 470.13 | 6.07 | 9.74 |
| 10000 | 5000×7500 | 53.33 | 366.6 | 9.47 |
| | 10000×15000 | 157.63 | 84.68 | 9.54 |
| | 15000×22500 | 309.2 | 35.47 | 9.62 |
| | 20000×30000 | 506.96 | 20.9 | 9.68 |
| 20000 | 5000×7500 m | 58.3 | 2532.12 | 9.26 |
| | 10000×15000 | 165.62 | 422.78 | 9.5 |
| | 15000×22500 | 327.15 | 171.19 | 9.57 |
| | 20000×30000 | 539.58 | 86.86 | 9.62 |

axis. In other words, the lower the network density, the lower the percentage of nodes having SF7 and higher percentage of nodes with SF12 to the gateway.



(a) 1000 nodes on $5000\text{m} \times 7500\text{m}$ area (b) 2500 nodes on $5000\text{m} \times 7500\text{m}$ area

Fig. 7: Maps of areas served by gateways selected by the proposed algorithm for two sample random topologies for redundancy factor $k = 1$.

5 Conclusion

We have proposed and evaluated an algorithm for sub-optimal selection of LPWAN gateway locations based on k -dominating set. The proposed algorithm provides the complete coverage for a given set of nodes, taking into account the maximum number of nodes served by a gateway and required redundancy. The performance evaluation using excessive simulations in both real-life topologies and regular topologies proved that the algorithm is able to select near-optimal locations with a low computing time.

References

1. Laiho J, Wacker A, Novosad T. Radio network planning and optimisation for UMTS. vol. 2. Wiley Online Library; 2002.
2. Hurley S. Planning effective cellular mobile radio networks. *IEEE Transactions on Vehicular Technology*. 2002;51(2):243–253.
3. Alber J, Fellows MR, Niedermeier R. Polynomial-time data reduction for dominating set. *Journal of the ACM (JACM)*. 2004;51(3):363–384.
4. CloudRF; https://cloudf.com/LoRa_planning.
5. Radio Mobile Online; <https://www.ve2dbe.com/rmonline.s.asp>.
6. Grochla K PK. Heuristic Algorithm for Gateway Location Selection in Large Scale LoRa Networks. In: *Proceedings of the 2020 International Conference on Wireless Communications & Mobile Computing (IWCMC 2020)*. IEEE; 2020.
7. Das B, Bharghavan V. Routing in ad-hoc networks using minimum connected dominating sets. In: *Communications, 1997. ICC'97 Montreal, Towards the Knowledge Millennium. 1997 IEEE International Conference on*. vol. 1. IEEE; 1997. p. 376–380.
8. Hochbaum DS, Shmoys DB. A best possible heuristic for the k -center problem. *Mathematics of operations research*. 1985;10(2):180–184.
9. Wan PJ, Alzoubi KM, Frieder O. Distributed construction of connected dominating set in wireless ad hoc networks. *Mobile Networks and Applications*. 2004;9(2):141–149.
10. Dai F, Wu J. On constructing k -connected k -dominating set in wireless networks. In: *Parallel and Distributed Processing Symposium, 2005. Proceedings. 19th IEEE International*. IEEE; 2005. p. 10–pp.
11. Foerster KT. Approximating fault-tolerant domination in general graphs. In: *Proceedings of the Meeting on Analytic Algorithmics and Combinatorics*. Society for Industrial and Applied Mathematics; 2013. p. 25–32.
12. Zhou T, Sun Y, He S, Shi Z, Chen J, Tao Z. Gateway Planning for Hybrid LoRa Networks. In: *2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. IEEE; 2019. p. 1071–1079.