

# Computational Complexity of Relay Placement in Sensor Networks\*

Jukka Suomela

Helsinki Institute for Information Technology HIIT,  
Basic Research Unit, Department of Computer Science,  
P.O. Box 68, FI-00014 University of Helsinki, Finland  
`jukka.suomela@cs.helsinki.fi`

**Abstract.** We study the computational complexity of relay placement in energy-constrained wireless sensor networks. The goal is to optimise balanced data gathering, where the utility function is a weighted sum of the minimum and average amounts of data collected from each sensor node. We define a number of classes of simplified relay placement problems, including a planar problem with a simple cost model for radio communication. We prove that all of these problem classes are NP-hard, and that in some cases even finding approximate solutions is NP-hard.

## 1 Introduction

In this article, we study the problem of placing relay nodes in *wireless sensor networks*. Sensor networks [1]-[4] consist of a large number of sensor nodes which collect data. The collected data is routed via the network to a sink node. The nodes are battery powered, and when considering battery lifetime, radio communication is a key issue [5].

Falck *et al.* [6] formulate the problem of *balanced data gathering* in sensor networks. In this formulation, the utility function is a weighted sum of the minimum and average amounts of data gathered from the nodes before the batteries are drained. The goal is to collect a large total amount of data, but not at the cost of completely ignoring some parts of the monitored area. Falck *et al.* show that the problem of finding an optimal routing can be presented as a linear program.

If the optimum is not satisfactory, one solution could be to add a small number of new *relay nodes* to the network. The obvious question is how to determine the optimal locations of the relays. This is the *relay placement problem*. Falck *et al.* [6] consider this problem briefly in the context of balanced data gathering. However, the computational complexity of this problem has not yet been analysed.

We will formalise the relay placement problem in Section 2. We will define various special cases or simplified versions of the general relay placement problem.

---

\* This work was supported in part by the Academy of Finland, Grant 202203, and by the IST Programme of the European Community, under the PASCAL Network of Excellence, IST-2002-506778. This publication only reflects the author's views.

We will then show in Section 3 that even these simplified versions are NP-hard, and we will show in Section 4 that in some cases even finding approximate solutions is NP-hard. Section 5 concludes this article.

## 2 Definitions of the Relay Placement Problems

An instance of the *balanced data gathering problem* [6], [7] is a tuple  $B = (\lambda, S, R, \sigma, E, s, \tau, \rho)$ . Here  $\lambda \in [0, 1]$  is a balance parameter,  $S$  is a finite set of sensor nodes,  $R$  is a finite set of relay nodes, and  $\sigma$  is the sink node. The sets  $S$ ,  $R$  and  $\{\sigma\}$  are disjoint. Let  $V = S \cup R \cup \{\sigma\}$ . The function  $E: V \rightarrow [0, \infty) \cup \{+\infty\}$  specifies the battery capacity of each node. The function  $s: S \rightarrow [0, \infty) \cup \{+\infty\}$  specifies how much data is available at each sensor node. The parameter  $\rho \in [0, \infty)$  is the cost of receiving one unit of data, and the function  $\tau: V \times V \rightarrow [0, \infty) \cup \{+\infty\}$  maps a pair of nodes to the cost of sending one unit of data from the first node to the second one. The solution to the problem is a flow  $f$ , where  $f_{\eta\kappa}$  is the amount of data transmitted from node  $\eta \in V$  to node  $\kappa \in V$ . The value  $q_\eta$  denotes the amount of data gathered from a node  $\eta \in S$ . The utility of the flow is  $\lambda \min_{\eta \in S} q_\eta + (1 - \lambda) \text{avg}_{\eta \in S} q_\eta$ .

An instance of the *relay placement problem* is a tuple  $P = (\lambda, S, \mathcal{R}, \sigma, E, s, \tau, \rho)$ ; the set of all such tuples is  $\mathcal{P}$ . Here  $\mathcal{R}$  is the set of *possible* relays, and the other parameters are as above. The sets  $S$ ,  $\mathcal{R}$ , and  $\{\sigma\}$  are disjoint. Let  $\mathcal{V} = S \cup \mathcal{R} \cup \{\sigma\}$ . The battery capacity function  $E(\eta)$  must be defined for all possible nodes  $\eta \in \mathcal{V}$ , and the transmission cost function  $\tau(\eta, \kappa)$  must be defined for all pairs of possible nodes  $\eta, \kappa \in \mathcal{V}$ . We will also assume that the location of the node,  $l(\eta) \in \mathbb{R}^2$ , is defined for all  $\eta \in \mathcal{V}$ . The solution is a finite subset  $R$  of possible relays  $\mathcal{R}$ . Given a relay placement instance  $P$  and its solution  $R$ , we can define the corresponding balanced data gathering instance  $B = (\lambda, S, R, \sigma, E|_V, s, \tau|_{V \times V}, \rho)$ , where  $V = S \cup R \cup \{\sigma\}$ . The utility of this solution,  $U(P, R)$ , is the maximum utility of  $B$ .

An instance of the *decision problem* is a tuple  $(P, N, u)$  where  $P \in \mathcal{P}$ ,  $N$  is the number of relays, and  $u$  is the utility requirement. The answer to the decision problem is *yes* if and only if there is a solution  $R$  to the relay placement problem  $P$  such that  $|R| = N$  and  $U(P, R) \geq u$ .

An instance of the *relay-constrained problem* is a pair  $(P, N)$  where  $P \in \mathcal{P}$  and  $N$  is the number of relays. The solution is any  $R \in \mathcal{R}$  with  $|R| = N$ . A solution  $R^*$  is optimal if it maximises  $U(P, R^*)$ . A solution  $\tilde{R}$  is  $k$ -optimal if  $U(P, \tilde{R}) \geq \frac{1}{k} U(P, R^*)$ .

An instance of the *utility-constrained problem* is a pair  $(P, u)$  where  $P \in \mathcal{P}$  and  $u$  is the utility requirement. The solution is any  $R \in \mathcal{R}$  with  $U(P, R) \geq u$ . A solution  $R^*$  is optimal if it minimises  $|R^*|$ . A solution  $\tilde{R}$  is  $k$ -optimal if  $|\tilde{R}| \leq k|R^*|$ .

A problem instance  $P \in \mathcal{P}$  is *planar*, denoted by  $P \in \mathcal{P}_P$ , if the set of possible relays  $\mathcal{R}$  is the plane  $\mathbb{R}^2$ , and  $l(\eta) = \eta$  for all  $\eta \in \mathcal{R}$ . A problem instance  $P \in \mathcal{P}$  has a *finite relay set*, denoted by  $P \in \mathcal{P}_D$ , if  $\mathcal{R}$  is finite. A problem instance  $P \in \mathcal{P}$  uses the *sensor upgrade model*, denoted by  $P \in \mathcal{P}_U$ , if  $\mathcal{R} = l(S)$ . Note that  $\mathcal{P}_U \subseteq \mathcal{P}_D$ .

A problem instance  $P \in \mathcal{P}$  has *location-dependent* transmission costs, denoted by  $P \in \mathcal{P}_L$ , if  $\tau(\eta, \kappa) = \tau'(l(\eta), l(\kappa))$  for some function  $\tau'$ . A problem instance  $P \in \mathcal{P}_L$  uses the *line-of-sight model*, denoted by  $P \in \mathcal{P}_S$ , if transmission costs can be defined by some parameters  $\alpha, p$ , and  $O$  as follows: The finite set  $O$  consists of disjoint obstacles; each obstacle is a simple (i.e., not self-intersecting) polygon in the real plane. The transmission cost  $\tau'(l_1, l_2)$  is infinite if the line segment  $\overline{l_1 l_2}$  intersects some obstacle  $o \in O$ . Otherwise,  $\tau'(l_1, l_2) = d_p(l_1, l_2)^\alpha$  where  $d_p(\cdot, \cdot)$  denotes the distance measured using the  $p$ -norm<sup>1</sup>. A problem instance  $P \in \mathcal{P}_S$  uses the *free space model*, denoted by  $P \in \mathcal{P}_F$ , if  $O = \emptyset$ .

A problem instance  $P \in \mathcal{P}$  has *identical batteries*, denoted by  $P \in \mathcal{P}_I$ , if there is  $E$  such that  $E(\eta) = E$  for all possible relays  $\eta \in \mathcal{R}$ .

We will denote  $\mathcal{P}_x \cap \mathcal{P}_y$  by  $\mathcal{P}_{xy}$ , etc. One can construct a total of 32 relay placement problem classes:  $\mathcal{P}, \mathcal{P}_P, \mathcal{P}_D, \dots, \mathcal{P}_{UFI}$ . We will denote the set of these classes by  $\mathcal{P}^*$  and we will use  $\mathcal{P}_x$  to refer an arbitrary member of  $\mathcal{P}^*$ .

### 3 All Problem Classes Are NP-Hard

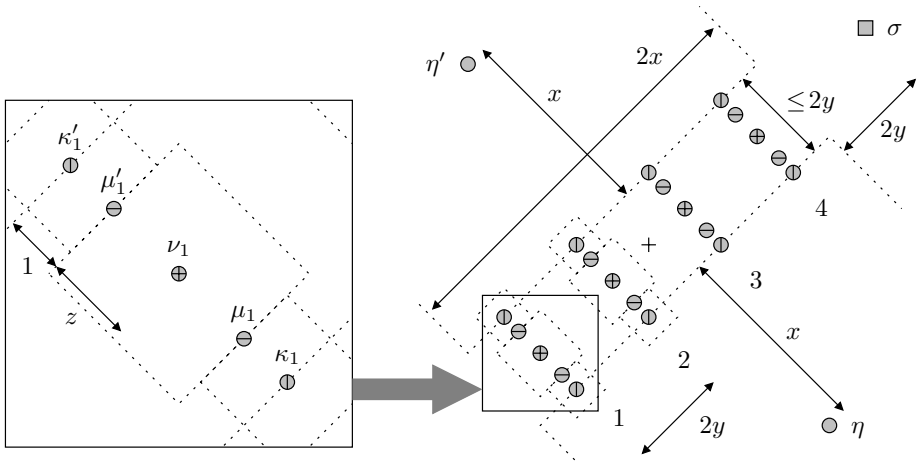
PARTITION is a well-known NP-complete problem [13], [14]. An instance of the PARTITION problem consists of a list of positive integers,  $(a_1, \dots, a_n)$ . A set  $X \subseteq \{1, 2, \dots, n\}$  is a feasible solution if  $\sum_{i \in X} a_i = \sum_{i \notin X} a_i$ . We will develop a polynomial reduction from PARTITION to  $\mathcal{P}_{UFI}$  and  $\mathcal{P}_{PFI}$ .

Let a list of positive integers,  $(a_1, \dots, a_n)$ , be given. We will assume that the sum of the integers is even; otherwise the answer to the problem would be trivially *no*. Construct a relay placement problem instance  $P$  as follows. First, define  $a^* = \max a_i$ , and  $b = \frac{1}{2} \sum a_i$ . Choose  $\lambda = 0, p = 1, \alpha = 2$ , and  $\rho = 0$ . Choose any values  $z \geq (na^*)^{1/\alpha}, y \geq z + 1$ , and  $x \geq ny$ .

Construct the problem geometry as shown in Fig. 1. Firstly, there are 2 sensors,  $\eta$  and  $\eta'$ , with  $E(\eta) = E(\eta') = bx^\alpha, s(\eta) = s(\eta') = b, l(\eta) = (z/2 + 1/2 + x/2, -z/2 - 1/2 - x/2)$ , and  $l(\eta') = -l(\eta)$ . Then, there are  $n$  diagonal rows of nodes, each row corresponding to one integer in the PARTITION problem. The centre points of these rows are  $l_i = ((2i - n - 1)y/2, (2i - n - 1)y/2)$ . On each row, there are two sensors,  $\kappa_i$  and  $\kappa'_i$ , with  $E(\kappa_i) = E(\kappa'_i) = a_i, s(\kappa_i) = s(\kappa'_i) = 0, l(\kappa_i) = l_i + (z/2 + 1/2, -z/2 - 1/2)$ , and  $l(\kappa'_i) = l_i - (z/2 + 1/2, -z/2 - 1/2)$ . Furthermore, on each row there are two sensors,  $\mu_i$  and  $\mu'_i$ , with  $E(\mu_i) = E(\mu'_i) = 0, s(\mu_i) = s(\mu'_i) = 0, l(\mu_i) = l_i + (z/2, -z/2)$ , and  $l(\mu'_i) = l_i - (z/2, -z/2)$ . The only purpose of these nodes is to act as possible relay locations in the sensor upgrade model. Finally, on each row there is one sensor,  $\nu_i$ , with  $E(\nu_i) = z^\alpha, s(\nu_i) = 1$ , and  $l(\nu_i) = l_i$ . The location of the sink is  $l(\sigma) = (x/2 + y, x/2 + y)$  and the battery capacity of the sink is irrelevant as the reception cost is zero. For the battery capacity of the relays, choose any value  $E \geq (a^* + 1)(2x + 2y)^\alpha$ .

The total number of sensor nodes is  $m = 5n + 2$ , and the total amount of available data is  $2b + n$  units. The utility of any solution is thus at most

<sup>1</sup> This simple power law has both theoretical [8] and practical [9]-[11] foundations, and non-Euclidean distance metrics may be a useful approximation in certain urban environments [12].



**Fig. 1.** Reduction from PARTITION to  $\mathcal{P}_{UFI}$  and  $\mathcal{P}_{PFI}$ . In this example, the corresponding PARTITION problem instance consisted of four integers. The diagonal rows labelled with numbers 1–4 correspond to the four integers.

$U^* = (2b + n)/m$ . We can now formulate the following decision problem instance:  $P$  is the relay placement problem instance constructed above, the number of relays  $N$  is  $n$ , and the utility requirement  $u$  equals  $U^*$ . We will show that this formulation is indeed a polynomial reduction from PARTITION to  $\mathcal{P}_{UFI}$  and  $\mathcal{P}_{PFI}$ .

**Lemma 1.** *Constructing the problem instance is possible in polynomial time.*

*Proof.* We may choose  $z = na^*$ ,  $y = z + 1$ ,  $x = ny$ , and  $E = (a^* + 1)(2x + 2y)^\alpha$ . The total number of nodes in the constructed problem is  $O(n)$ . The parameters of each node can be calculated in polynomial time: keeping in mind that  $\alpha = 2$ , all expressions above only involve integer or rational numbers, and the size of each integer is polynomial in the size of the input.

**Lemma 2.** *If the answer to the PARTITION problem instance is yes, the answer to the relay placement problem instance constructed above is yes, both in the  $\mathcal{P}_{UFI}$  and in the  $\mathcal{P}_{PFI}$  formulation.*

*Proof.* Let  $X$  be a feasible solution to the PARTITION problem. Denote the set  $\{1, \dots, n\} \setminus X$  by  $X'$ . For each  $i \in X$ , place a relay on  $\mu_i$ , transmit  $a_i$  units of data from  $\eta$  to  $\kappa_i$ , forward  $a_i$  units of data from  $\kappa_i$  to the relay at  $\mu_i$ , transmit 1 unit of data from  $\nu_i$  to the relay at  $\mu_i$ , and forward  $a_i + 1$  units of data from the relay at  $\mu_i$  to the sink. For each  $i \in X'$ , place a relay on  $\mu'_i$ , and construct flows as above for  $\eta'$  and  $\kappa'_i$ .

**Lemma 3.** *If the answer to the PARTITION problem instance is no, the answer to the relay placement problem instance constructed above is no, both in the  $\mathcal{P}_{UFI}$  and in the  $\mathcal{P}_{PFI}$  formulation.*

*Proof.* Let us assume that the answer to the relay placement problem instance is *yes*. This is possible only if all available data from all sensor nodes is forwarded to the sink node.

Let us first study the node  $\nu_i$  for some  $i$ . Due to its limited battery capacity, the sensor has to send at least some of its data to a node whose distance is at most  $z$  units. No sink or sensor node with a positive battery capacity is available within the area of distance  $z$  from  $\nu_i$ . Thus, at least one relay node has to be located in this area. As there are  $n$  such areas, all non-overlapping, there must be exactly one relay node in each area.

Let us denote by  $X$  the indexes of the areas where the relay is closer to  $\eta$  than to  $\eta'$ . Denote  $\{1, \dots, n\} \setminus X$  by  $X'$ . As the answer to the PARTITION problem was *no*,  $\sum_{i \in X} a_i \neq \sum_{i \in X'} a_i$ . Without loss of generality, we assume that  $\sum_{i \in X} a_i < \sum_{i \in X'} a_i$ . Clearly  $\sum_{i \in X} a_i < b$ . As  $b$  and  $a_i$  are integral,  $\sum_{i \in X} a_i \leq b - 1$ .

The sensor  $\eta$  has to send  $b$  units of data to other nodes. The node has enough energy resources for transmitting  $b$  units of data to the distance of exactly  $x$  units. If some part of the data was sent over a larger distance, another part would have to be sent to a node whose distance is less than  $x$  units; however, no sensor or sink node is available closer than this, and all relays are already tied to the proximity of nodes  $\nu_i$ . Thus, the only possibility is to send all data to nodes  $\kappa_i$ , each exactly  $x$  units from the source node. Let the amount of data transmitted from  $\eta$  to  $\kappa_i$  be  $c_i$ . Clearly  $\sum c_i = b$  and  $c_i \geq 0$ .

Now,  $\sum_{i \in X} a_i \leq b - 1 = \sum (c_i - 1/n)$ . At least one of the following holds: there is  $i \in X$  such that  $a_i \leq c_i - 1/n$ , or there is  $i \in X'$  such that  $c_i \geq 1/n$ . If neither holds, then  $\sum_{i \in X} a_i > \sum_{i \in X} (c_i - 1/n) \geq \sum (c_i - 1/n)$ , a contradiction.

Let us first assume that there is  $i$  such that  $i \in X$  and  $a_i \leq c_i - 1/n$ . In this case the node  $\kappa_i$  would have to transmit at least  $a_i + 1/n$  units of data to some other node. The distance to the closest node is at least 1 unit. Thus, the transmission cost is at least  $a_i + 1/n$ , exceeding the available battery capacity  $a_i$ , a contradiction.

On the other hand, if there is  $i$  such that  $i \in X'$  and  $c_i \geq 1/n$ , the node  $\kappa_i$  would have to transmit at least  $1/n$  units of data to some other node. As  $i \in X'$ , the distance to the closest relay node is at least  $z + 1$  units. The only node less than  $z + 1$  units from  $\kappa_i$  is  $\nu_i$ , and it does not have any battery capacity for forwarding data. Thus, we need to transmit at least  $1/n$  units of data to a distance of at least  $z + 1$  units, requiring at least  $(1/n)(z + 1)^\alpha$  units of energy. Here  $(1/n)(z + 1)^\alpha \geq (1/n)((na^*)^{1/\alpha} + 1)^\alpha > (1/n)((na^*)^{1/\alpha})^\alpha = (1/n)(na^*) = a^* \geq a_i = E(\kappa_i)$ . Again, a contradiction. Thus the assumption must be false.

**Theorem 1.** *The decision versions of all relay placement problem classes in  $\mathcal{P}^*$  are NP-hard.*

*Proof.* From the list of problem definitions in Section 2 we see that for any relay problem class  $\mathcal{P}_x$  in  $\mathcal{P}^*$ , either  $\mathcal{P}_{UFI} \subseteq \mathcal{P}_x$  or  $\mathcal{P}_{PFI} \subseteq \mathcal{P}_x$ . The theorem follows from Lemmas 1, 2, and 3.

**Theorem 2.** *The decision version of the relay placement problem class  $\mathcal{P}_D$  is NP-complete.*

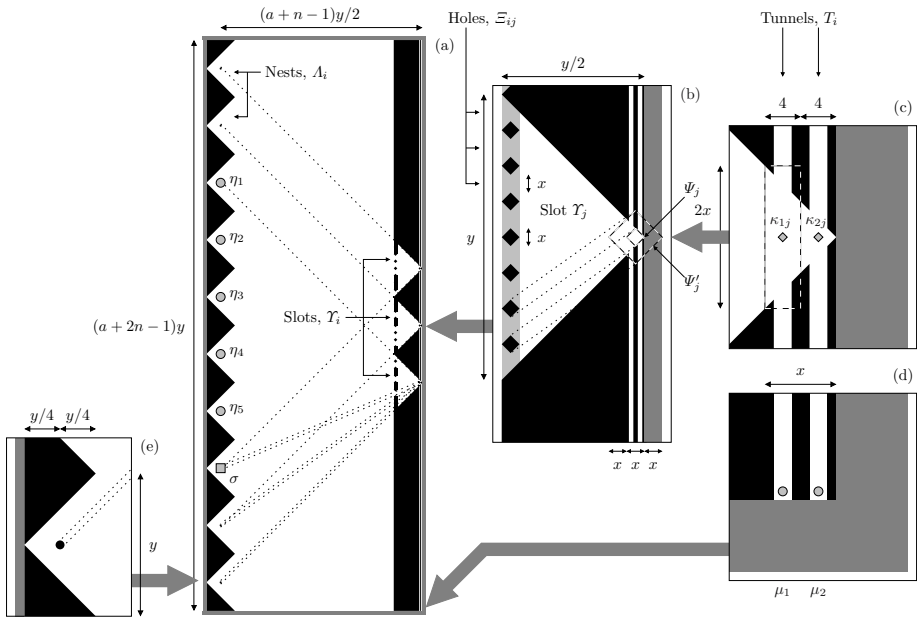
*Proof.* Use  $(R, f)$  as a certificate for a *yes* instance of the decision problem.

### 4 For Some Problem Classes, Approximation Is NP-Hard

SET COVERING is another well-known NP-complete problem [13,14]. An instance of the SET COVERING problem consists of a finite collection of finite sets,  $\mathcal{A} = \{A_1, \dots, A_n\}$ , and a positive integer  $m$ . A subcollection  $\mathcal{X} \subseteq \mathcal{A}$  is a feasible solution if  $|\mathcal{X}| \leq m$  and  $\bigcup \mathcal{X} = \bigcup \mathcal{A}$ . We will now develop a polynomial reduction from SET COVERING to  $\mathcal{P}_{DSI}$  and  $\mathcal{P}_{PSI}$ .

Let  $\mathcal{A} = \{A_1, \dots, A_n\}$  and  $m$  be given. Let  $a = |\bigcup \mathcal{A}|$ . Without loss of generality we will assume that  $\bigcup \mathcal{A} = \{1, \dots, a\}$ . Construct a relay placement problem instance  $P$  as follows. Choose  $\lambda = 1$ ,  $p = 2$ ,  $\alpha = 2$ , and  $\rho = 0$ . Define  $x = 4m$  and  $y = 2x(a + n)$ . Construct the problem geometry as shown in Fig. 2.

On the left-hand side of part (a), we have  $a + 2n - 1$  triangular *nests*. The first  $n - 1$  nests are empty. The next  $a$  nests,  $A_1$  to  $A_a$ , correspond to integers  $\{1, \dots, a\}$ , where the nest  $A_i$  contains the sensor node  $\eta_i$ , with  $E(\eta_i) = 1$ ,  $s(\eta_i) = 1$ . The next nest,  $A_\sigma$ , contains the sink node  $\sigma$ . The last  $n - 1$  nests are empty.



**Fig. 2.** Reduction from SET COVERING to  $\mathcal{P}_{DSI}$  and  $\mathcal{P}_{PSI}$ . This figure illustrates the case of  $m = 2$ ,  $n = 3$ , and  $a = 5$ . Some details are shown in a larger scale.

On the right-hand side, we have  $n$  triangular *slots*,  $\mathcal{Y}_1$  to  $\mathcal{Y}_n$ . Each slot corresponds to one element of  $\mathcal{A}$ . Let us now have a closer look at one of these slots, let it be slot  $\mathcal{Y}_j$ . See Fig. 2 (b) for an illustration. On the leftmost side of the slot, we have  $a + n - 1$  diamond-shaped obstacles. Between the diamond-shaped obstacles, we have  $a + n$  *holes*. The first  $n - j$  holes are unused. The next  $a$  holes,  $\Xi_{1j}$  to  $\Xi_{aj}$ , correspond to the sensors  $\eta_1$  to  $\eta_a$ , and the next hole,  $\Xi_{\sigma j}$  corresponds to the sink  $\sigma$ . Finally, there are  $j - 1$  unused holes.

Let us now construct two diamond-shaped areas,  $\Psi_j$  and  $\Psi'_j$ , as illustrated in Fig. 2 (b). All points  $l \in \Psi_j$  satisfy the following conditions: for all  $i$ , there is a line of sight from  $l$  to  $\eta_i$  through  $\Xi_{ij}$ ; and there is a line of sight from  $l$  to  $\sigma$  through  $\Xi_{\sigma j}$ . All points  $l \in \Psi'_j$  satisfy the following conditions: for all  $i$ , if there is a line of sight from  $l$  to  $\eta_i$ , it necessarily passes through  $\Xi_{ij}$ ; and if there is a line of sight from  $l$  to  $\sigma$ , it necessarily passes through  $\Xi_{\sigma j}$ .

Now, we will block the hole  $\Xi_{ij}$  if and only if  $i \notin A_j$ . Let us denote by  $X_l$  the set of indexes  $j$  such that  $\eta_j$  is still visible from the location  $l$ . We can make two observations: If  $l \in \Psi_j$ , then  $X_l = A_j$ . If  $l \in \Psi'_j$ , then  $X_l \subseteq A_j$ .

We will also need  $m$  narrow, vertical tunnels,  $T_1$  to  $T_m$ , in the rightmost part of the construction; see parts (c) and (d) for an illustration. Each tunnel consists of a 1-unit-wide wall, a 2-unit-wide tunnel, and a 1-unit-wide wall, and we will refer to the *interior* of this 4-unit-wide area as  $T'_i$ . For each  $i$ , place a sensor node  $\mu_i$  at the bottom of tunnel  $T_i$ , with  $E(\mu_i) = 1$  and  $s(\mu_i) = 1$ . Note that all points in  $T_i$  are visible from  $\mu_i$ , and no point outside  $T'_i$  is visible from  $\mu_i$ .

At the intersection of the tunnel  $T_i$  and the slot  $\mathcal{Y}_j$  there is a possible relay location  $\kappa_{ij}$ . Note that this location is inside area  $\Psi_j$ . Finally, the construction is surrounded by four walls, shown in the figure in grey colour.

All relays have a battery capacity of 1 unit. We can now formulate the following relay-constrained optimisation problem instance:  $P$  is the relay placement problem instance constructed above, and the number of relays  $N$  is  $m$ .

**Lemma 4.** *Constructing this relay placement problem instance is possible in polynomial time.*

*Proof.* The construction involves generating a problem instance with  $O(a + n)$  sensors,  $O(nm)$  possible relays, and  $O((a+n)n)$  quadrilateral or triangular obstacles. Calculating the parameters of each node and each obstacle can be performed in a polynomial time. The calculations only involve integers.

**Lemma 5.** *If the answer to the SET COVERING problem instance is yes, the optimal solution to the relay placement problem instance constructed above has a positive utility, both in the  $\mathcal{P}_{DSI}$  and in the  $\mathcal{P}_{PSI}$  formulation.*

*Proof.* Let  $\mathcal{X}' = \{A_{c_1}, \dots, A_{c_{m'}}\}$  be a solution to the SET COVERING problem, with  $m' \leq m$ . Choose, for example,  $c_i = 1$  for all  $i > m'$ . Now  $\mathcal{X} = \{A_{c_1}, \dots, A_{c_m}\}$  is still a feasible solution. For each  $i \in \{1, \dots, m\}$ , place a relay at  $\kappa_{ic_i}$ . This way all sensors can forward some data through relays.

**Lemma 6.** *If the answer to the SET COVERING problem instance is no, there is no solution with a positive utility, either in  $\mathcal{P}_{DSI}$  or in  $\mathcal{P}_{PSI}$ .*

*Proof.* If the utility is positive, we have to gather some data from all sensors. For each  $i$ , there has to be at least one relay on  $T'_i$ , let us call it  $\nu_i$ . Thus, all  $m$  relays are bound to tunnels. Let  $Y_i = X_{l(\nu_i)}$ , and  $\mathcal{Y} = \{Y_1, \dots, Y_m\}$ . To gather data from each  $\eta_j$ , we must have  $\bigcup \mathcal{Y} = \bigcup \mathcal{A}$ . If the relay  $\nu_i$  is located in some  $\Psi'_j$ , we choose  $c_i = j$ ; otherwise the relay must be inside a tunnel,  $Y_i$  is  $\emptyset$ , and we can choose  $c_i = 1$ . For each  $i$  there is now a  $c_i$  such that  $Y_i \subseteq A_{c_i}$ . Define  $\mathcal{Y}' = \{A_{c_1}, \dots, A_{c_m}\}$ . Now we have  $\bigcup \mathcal{A} = \bigcup \mathcal{Y} \subseteq \bigcup \mathcal{Y}' \subseteq \bigcup \mathcal{A}$ . Thus,  $\mathcal{Y}'$  is a feasible solution to the SET COVERING problem instance.

**Theorem 3.** *Finding  $k$ -optimal solutions to the relay-constrained optimisation versions of problem classes  $\mathcal{P}_x$  satisfying  $\mathcal{P}_{DSI} \subseteq \mathcal{P}_x$  or  $\mathcal{P}_{PSI} \subseteq \mathcal{P}_x$  is NP-hard.*

*Proof.* Let us assume that for some  $k$ , we have an oracle for solving the relay-constrained optimisation problems of class  $\mathcal{P}_{DSI}$  or  $\mathcal{P}_{PSI}$   $k$ -optimally in constant time. We may then use the construction presented above to solve SET COVERING in polynomial time.

By Lemma 4, we may construct the relay placement problem instance in polynomial time. By Lemmas 5 and 6, the oracle will return a solution with a positive utility if and only if the answer to the SET COVERING problem is *yes*.

## 5 Conclusions

In this article, we have specified and studied a number of classes of relay placement problems. All classes have been proved NP-hard. For some important problem classes, approximation has been proved NP-hard as well. It is an open question whether it is possible to formulate a relay placement problem which is computationally tractable but still meaningful in practise. We may need to consider other utility functions in addition to the balanced data gathering formulation.

However, these results do not prevent us from optimising relay placement. One possibility is to use a heuristic approach [6], which is computationally effective. While it does not guarantee optimality, it may still be useful in practical problems. There are also algorithms for finding  $k$ -optimal solutions [15]. The time complexity of these algorithms may be high, but they have been successfully used for solving problem instances of moderate size.

**Acknowledgements.** Thanks to Patrik Floréen, Jyrki Kivinen, Tanja Säily, Jukka Kohonen, and Petteri Nurmi for discussions and comments.

## References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., and Cayirci, E.: Wireless Sensor Networks: a Survey. *Computer Networks* **38** (2002) 392–422
2. Al-Karaki, J.N., and Kamal, A.E.: Routing Techniques in Wireless Sensor Networks: a Survey. *IEEE Wireless Communications* **11** (2004) 6–28
3. Culler, D., Estrin, D., and Srivastava, M.: Guest Editors' Introduction: Overview of Sensor Networks. *IEEE Computer* **37** (2004) 41–49



4. Ephremides, A.: Energy Concerns in Wireless Networks. *IEEE Wireless Communications* **9** (2002) 48–59
5. Raghunathan, V., Schurgers, C., Park, S., and Srivastava, M.B.: Energy-Aware Wireless Microsensor Networks. *IEEE Signal Processing Magazine* **19** (2002) 40–50
6. Falck, E., Floréen, P., Kaski, P., Kohonen, J., and Orponen, P.: Balanced Data Gathering in Energy-Constrained Sensor Networks. In: *Algorithmic Aspects of Wireless Sensor Networks: First International Workshop (ALGOSENSORS 2004, Turku, Finland, July 2004)*, Berlin Heidelberg, Springer-Verlag (2004) 59–70
7. Floréen, P., Kaski, P., Kohonen, J., and Orponen, P.: Exact and Approximate Balanced Data Gathering in Energy-Constrained Sensor Networks. *Theoretical Computer Science* (2005) To appear.
8. Rappaport, T.S.: *Wireless Communications, Principles and Practice*. Prentice Hall, Inc., Upper Saddle River (1999)
9. Andersen, J.B., Rappaport, T.S., and Yoshida, S.: Propagation Measurements and Models for Wireless Communications Channels. *IEEE Communications Magazine* **33** (1995) 42–49
10. Seidel, S.Y., and Rappaport, T.S.: 914 MHz Path Loss Prediction Models for Indoor Wireless Communications in Multifloored Buildings. *IEEE Transactions on Antennas and Propagation* **40** (1992) 207–217
11. Sohrabi, K., Manriquez, B., and Pottie, G.J.: Near Ground Wideband Channel Measurement in 800–1000 MHz. In: *Proceedings of the 49th Vehicular Technology Conference* (1999) 571–574
12. Goldsmith, A.J., and Greenstein, L.J.: A Measurement-Based Model for Predicting Coverage Areas of Urban Microcells. *IEEE Journal on Selected Areas in Communications* **11** (1993) 1013–1023
13. Karp, R.M.: Reducibility among Combinatorial Problems. In Miller, R.E., Thatcher, J.W., eds.: *Complexity of Computer Computations*, Plenum Press, New York (1972) 85–103
14. Garey, M.R., and Johnson, D.S.: *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, New York (2003)
15. Suomela, J.: Algorithms for  $k$ -Optimal Relay Placement in Sensor Networks (2005) Submitted for publication.