Approximating 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

The problem of placing relay nodes in a wireless sensor network is studied in the context of balanced data gathering. Previous work is extended by showing that even the simplest classes of the relay placement problem are hard to approximate. This work also presents a heuristic method for both lower-bounding and upper-bounding the maximum performance of a sensor network over all possible relay locations.

Categories and Subject Descriptors: C.2.1 [Computer-communication Networks]: Network Architecture and Design; F.2.2 [Analysis of Algorithms and Problem Complexity]: Nonnumerical Algorithms and Problems

General Terms: Performance, Algorithms, Theory

Keywords: wireless sensor networks, relay placement, balanced data gathering, energy constraints

1. INTRODUCTION

Wireless sensor networks [5] 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, one of the key issues is radio communication [6]. To increase the amount of data gathered from a sensor network during its lifetime, one may add a small number of relay nodes (base stations) that are equipped with large batteries. The relay nodes may gather data from nearby sensor nodes and forward it towards the sink node.

This work focuses on the *relay placement problem* [7]: given a sensor network and the number of relay nodes, the goal is to find optimal locations for the relay nodes. The performance of a network is determined by the amount of data gathered from each sensor node during the lifetime of the network, i.e., before the batteries are drained. The *balanced data gathering* [1, 3] formulation is used: the utility function is a weighted sum of the minimum and average amounts of data gathered from the nodes.

This paper is organised as follows. Section 2 defines the problem, including two special cases, the Euclidean problem and the sensor-upgrade problem. Section 3 shows that even these simple special cases are NP-hard to approximate within small constant factors. Section 4 presents a heuristic algorithm for both lower-bounding and upper-bounding the Euclidean relay placement problem.

2. BACKGROUND

An instance of the *balanced data gathering problem* [1, 3] is a tuple $(\lambda, S, R, \sigma, E, s, \tau, \rho)$, where $\lambda \in [0, 1]$ is a balance parameter, S is a finite set of sensor nodes, R is a finite set of relay nodes, σ is the sink node, E_i specifies the battery capacity of the node i, s_i specifies how much data is available at the sensor node i, τ_{ij} is the cost of sending one unit of data from i to j, and ρ is the cost of receiving one unit of data. The solution of the problem is a flow f, where f_{ij} is the nonnegative amount of data transmitted from i to j. The nonnegative value $q_i \le s_i$ denotes the amount of data gathered from the node $i \in S$. The flow must be preserved: $q_i + \sum_j f_{ji} = \sum_j f_{ij}$ for each sensor i and $\sum_j f_{ji} = \sum_j f_{ij}$ for each relay i. Furthermore, transmission and reception costs must not exceed the battery capacity: $\sum_j \tau_{ij} f_{ij} + \sum_j \rho f_{ji} \le E_i$ for each node i. The utility of the solution is the weighted sum of the minimum and average amounts of data gathered, $\lambda \min_{i \in S} q_i + (1 - \lambda) \operatorname{avg}_{i \in S} q_i$.

An instance of the relay-constrained *relay placement problem* [7] is a tuple $(\lambda, S, \mathcal{R}, N, \sigma, E, s, \tau, \rho)$, where $\mathcal{R}$ is the set of *possible* relays, N is the number of relays, and the other parameters are as above in the balanced data gathering problem. The solution of the problem is a relay placement $R \subseteq \mathcal{R}$ with $|R| = N$. Given a solution R, one can construct the corresponding instance of the balanced data gathering problem, $(\lambda, S, R, \sigma, E, s, \tau, \rho)$. The utility of the solution R is the maximum of the utility of the balanced data gathering problem. A solution is h-approximate if its utility is at least $1/h$ times the optimum.

In this work, it is assumed that the battery capacity E_i is the same for all possible relays $i \in \mathcal{R}$.

The transmission cost τ_{ij} must be defined for all pairs of possible nodes, $i, j \in S \cup \mathcal{R} \cup \{\sigma\}$. These can be specified explicitly in the *finite problem*, where $\mathcal{R}$ is a finite set. However, this is not possible if $\mathcal{R}$ is infinite; thus, the main focus is on the following simple model for transmission costs.

Each node is associated with a location in the Euclidean plane $\mathbb{R}^2$. The transmission costs are defined by $\tau_{ij} = d(i, j)^\alpha$ where $\alpha > 0$ is a parameter (typically $2 \le \alpha \le 4$) and $d(\cdot, \cdot)$ denotes the Euclidean distance. This model is used in the following problems: In the *Euclidean problem*,

			$\alpha = 2.0$	$\alpha = 3.0$	$\alpha = 4.0$
(a)	$\lambda = 1.0$	$\rho > 0$	3.24	5.85	10.56
		$\rho = 0$	2.99	5.19	8.99
	$\lambda = 0.5$	$\rho > 0$	1.52	1.70	1.82
		$\rho = 0$	1.49	1.67	1.79
(b)	$\lambda = 1.0$	$\rho > 0$	3.99	7.99	15.99
		$\rho = 0$	1.14	1.14	1.14
	$\lambda = 0.5$	$\rho > 0$	1.59	1.77	1.88
		$\rho = 0$	1.06	1.06	1.06

Table 1: Inapproximability ratios for the relay placement problem: (a) Euclidean problem, (b) sensor-upgrade problem.

the set of possible relays is the plane $\mathbb{R}^2$. In the *sensor-upgrade problem*, the set of possible relays equals the set of sensor locations.

Note that the sensor-upgrade problem is a special case of the finite problem. It would be possible to specify arbitrary transmission costs τ_{ij} explicitly. However, by using the above definition it is possible to show that the sensor-upgrade problem (and, thus, the finite problem) remains computationally hard even if the transmission costs are defined by a simple model of radio propagation.

Prior to this work, these special cases of the relay placement problem were known to be NP-hard to solve exactly [7]. However, it was not known whether they are also hard to approximate.

3. INAPPROXIMABILITY

This section shows that both the Euclidean problem and the sensor-upgrade problem are NP-hard to approximate within small constant factors. Even if the parameters of the model for transmission cost are fixed to physically realistic values, inapproximability within factors as high as 10 can be obtained. Table 1 summarises the key results for some concrete numerical values. For example, even if one fixes $\lambda = 1$, $\alpha = 3$, and an arbitrary $\rho > 0$, the above-mentioned simplified problems are NP-hard to approximate within a constant factor of 5 or better, i.e., any polynomial-time algorithm may return solutions with a utility as low as 20 % of the optimum, assuming P $\neq$ NP.

The proof of the inapproximability of relay placement uses the same idea as Feder and Greene [2] in their proof of the inapproximability of k-centre clustering.

In this section, the term *3-planar graph* refers to a planar graph with maximum degree 3, and the term *3-planar vertex covering* refers to the problem of determining whether there is a vertex cover of a given size for a given 3-planar graph. In this section, an instance of 3-planar vertex covering is first transformed to an equivalent *embedded* instance of 3-planar vertex covering; this is further reduced to the problem of *approximating* relay placement. This proves that approximating relay placement is NP-hard, as 3-planar vertex covering is NP-complete [4].

Transformation. Consider a 3-planar graph G_0 (Figure 1a). Remove isolated vertices, if any. Embed the graph in a rectangular two-dimensional grid of polynomial size (Figure 1b). This is possible in polynomial time; see, e.g., the linear-time algorithm by Tamassia and Tollis [8].

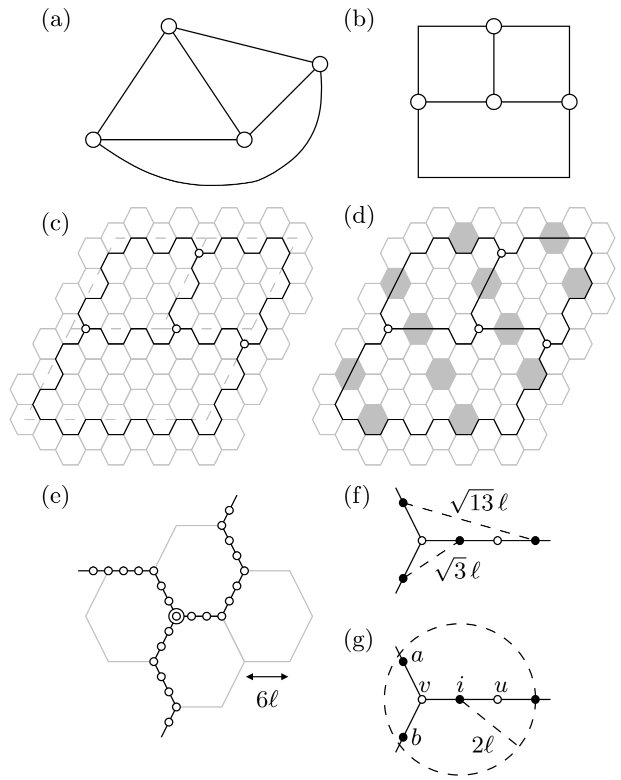

Figure 1: Transformation and some key properties.

Shear the rectangular grid and draw the edges along a regular hexagonal grid (Figure 1c). The sides of the hexagons are of length 6ℓ; the value $\ell > 0$ will be chosen later. In the hexagonal grid, there are tiles that correspond to the midpoints of the line segments of the original rectangular grid (greyed in Figure 1d); a shortcut can be taken through such tiles. For each edge, take one shortcut if needed to ensure that the length of each edge is an odd multiple of 6ℓ (Figure 1d).

Form a new graph $G = (V, E)$ by splitting the edges of G_0 into pieces of length 2ℓ (Figure 1e). Let k' denote the total number of new vertices created by the splits; there is an even number of new vertices on each edge of G_0. For any k_0, let $k = k_0 + k'/2$. Now, (G, k) and (G_0, k_0) are equivalent instances of 3-planar vertex covering. Except for a scaling by factor 2ℓ, this is also a concrete way of constructing the embedded graph described by Feder and Greene [2, Figure 2a].

The locations of the vertices have to be presented by rational numbers. Choose rational x coordinates. Scale down the height of the hexagons by less than ϵ units to find rational y coordinates. It is assumed that $0 < \epsilon < (\sqrt{13}/2 - \sqrt{3})\ell$.

Reduction. Vertices of degree 1 are called *outer vertices*; the remaining vertices are called *inner vertices*. Edges adjacent to at least one outer vertex are called *outer edges*; the remaining edges are called *inner edges*. Midpoints of these edges are called *outer midpoints* and *inner midpoints*, respectively.

Construct the following three instances of the relay placement problem. First, construct an instance of the sensor-upgrade problem; the instance is called I_1. Choose $N = k$; the values of λ, α, and ρ are discussed later. Place one sensor at each inner midpoint, outer midpoint, inner vertex, and outer vertex. These are called *inner edge sensors*, *outer edge sensors*, *inner vertex sensors*, and *outer vertex sensors*, and their battery capacities are ℓ^α, ℓ^α, $(2\ell)^\alpha$, and $(2\ell)^\alpha$ units, respectively. The amount of available data is 1 for each sensor. Place the sink node at any location not closer than 3ℓ units from any vertex or edge of G. The battery capacity of each relay is ∞ (or a suitable large constant).

Second, construct another instance of the sensor-upgrade problem, I_2. The construction of I_2 is identical to I_1 except that the battery capacities of inner edge sensors, outer edge sensors, inner vertex sensors, and outer vertex sensors are $(3/2)\ell^\alpha$, $2\ell^\alpha$, ℓ^α, and ℓ^α units, respectively.

Third, construct an instance of the Euclidean problem, I_3. The network layout and the parameters of I_3 are identical to those of I_1 except that there are no vertex sensors.

"Yes" instances. Assume that there is a vertex cover $X \subseteq V$ of size k for G. Construct a relay placement R by placing one relay at each vertex $v \in X$. The set R is a feasible solution of I_1, I_2, and I_3.

Consider I_1 or I_3. For each vertex or edge sensor, there is enough battery capacity to transmit 1 unit of data directly to the nearest relay, and the relay can forward the data to the sink. The amount of data gathered from each sensor is 1, and the utility of the solution equals 1 for all λ, α, and ρ.

Next, consider I_2. If $\rho = 0$, it is possible to gather 1 unit of data from each sensor as follows. Edge sensors transmit directly to the nearest relay. Outer vertex sensors without a relay transmit 1 unit of data to the neighbouring outer edge sensor. Inner vertex sensors without a relay choose two neighbouring edge sensors and transmit $1/2$ units of data to each of them. The utility of the solution equals 1 for all λ and α, assuming $\rho = 0$.

"No" instances, Euclidean problem. Consider a solution of I_3. A relay *r-covers* an edge if the distance between the relay and the midpoint of the edge is at most r. The distance between the midpoints of two non-neighbouring edges is more than $\sqrt{13}\,\ell - \epsilon$ units [2], see Figure 1f. The degree of the graph is bounded by 3, and there are no cliques of size 3. Thus, if a set of edges is $(\sqrt{13}\,\ell/2 - \epsilon)$-covered by a relay, the edges are mutually neighbouring, and they share a common endpoint. If each edge is $(\sqrt{13}\,\ell/2 - \epsilon)$-covered by some relay, the corresponding common endpoints form a vertex cover of size at most $N = k$.

Thus, if there is no vertex cover of size k, there is an edge sensor i such that there is no relay or sink within $\sqrt{13}\,\ell/2 - \epsilon$ units [2]. There may be two other inner edge sensors a and b at a distance between $\sqrt{3}\,\ell - \epsilon$ and $\sqrt{3}\,\ell$, see Figure 1g. The sensors i, a, and b are called *nearby sensors*; other sensors, relays, and the sink are called *distant targets*.

Assume that at least q units of data is gathered from the sensor i. Write y for the total amount of data transmitted from i to a and b; the remaining $q - y$ units are transmitted to distant targets. The nodes a and b have to receive at least a total of y units of data. The total energy resources of the nodes imply $(\sqrt{3}\,\ell - \epsilon)^\alpha y + (\sqrt{13}\,\ell/2 - \epsilon)^\alpha (q - y) \le \ell^\alpha$ and $\rho y \le 2\ell^\alpha$. If λ, ρ, and α are part of the problem instance, and if arbitrary values are allowed for these parameters, choose $\ell = 1$, $\lambda = 1$, $\rho = 0$, and $\alpha = \log_{\sqrt{3}-\epsilon} h$ for any $h > 1$. It follows that $q \le 1/h$, and the utility of the solution is at most $1/h$. Thus, distinguishing between a utility of 1 and a utility of $1/h$ in the relay placement problem makes it possible to solve 3-planar vertex covering; in other words, approximating relay placement within any constant factor h is NP-hard. However, this says little about the approximability of practical, physically realistic instances of the relay placement problem.

Now, assume that both α and ρ are fixed to arbitrary values. It turns out that the relay placement problem is hard to approximate even in this case. First, consider the case $\rho > 0$. For any $\epsilon_1 > 0$, there is an ℓ such that $2\ell^\alpha/\rho < \epsilon_1$, implying $y < \epsilon_1$. Thus, it is possible to obtain an upper bound for q that is arbitrarily close to $(2/\sqrt{13})^\alpha$. By choosing $\lambda = 1$, the bound implies the inapproximability of relay placement within $(\sqrt{13}/2)^\alpha - \epsilon_2$ for all $\epsilon_2 > 0$. Second, consider the case $\rho = 0$. By choosing, say, $\ell = 1$, it is possible to obtain an upper bound arbitrarily close to $(1/\sqrt{3})^\alpha$, implying inapproximability within $(\sqrt{3})^\alpha - \epsilon_2$ for all $\epsilon_2 > 0$.

Even if λ is fixed to any value $0 < \lambda \le 1$, the same idea can be applied to obtain an inapproximability ratio strictly larger than 1 by noting that $\min q_i$ is bounded by q and $\operatorname{avg} q_i$ is bounded by 1. The inapproximability ratios for some concrete values are summarised in Table 1.

"No" instances, sensor-upgrade problem. Consider a solution of the instance I_1 or I_2. A disk of radius $2\ell - \epsilon$ is required in order to cover non-neighbouring edges by a relay, see Figure 1g. Thus, if there is no vertex cover of size k, there is an edge sensor i such that there is no relay or sink within $2\ell - \epsilon$ units. There are two vertex sensors u and v at a distance between $\ell - \epsilon$ and ℓ, and in the case of an inner edge sensor, there may be two other inner edge sensors a and b at distances between $\sqrt{3}\ell - \epsilon$ and $\sqrt{3}\ell$. The sensors i, u, v, a, and b are called *nearby sensors*; other sensors, relays, and the sink are called *distant targets*. Choose the labels so that the degree of the vertex u is always 2; the degree of the vertex v may be 1, 2, or 3.

Again, assume that both α and ρ are fixed to arbitrary values. First, consider the case $\rho > 0$. Study the instance I_1. As above, choose a small $\ell > 0$; this way receptions are very expensive compared to battery capacities, only a negligible amount of data gathered from i can be forwarded by nearby sensors, and the utility is bounded by the amount of data that can be transmitted over a distance of $2\ell - \epsilon$ units. This implies inapproximability within $2^\alpha - \epsilon_2$ for all $\epsilon_2 > 0$.

Second, consider the case $\rho = 0$. Study the instance I_2. To simplify the discussion, the proof focuses on on the special case of $\alpha > \log_3 4 \approx 1.26$; this covers most physically realistic special cases. Assume that at least q units of data is gathered from each of i, u, and v. Consider the total amount of energy consumed by i, u, and v.

If the degree of v is 1, there is no other node except i and u within $3\ell - \epsilon$ units. For transmitting data from v to a distant target, there are different multihop paths to choose from: directly from v to a distant target, via i to a distant target, etc. However, any combination of such paths consumes at least a total of $3q(\ell - \epsilon)^\alpha$ units of energy in i, u, and v if q units of data needs to be transmitted. Similarly, any routing from i to a distant target consumes at least $2q(\ell - \epsilon)^\alpha$ units of energy, and any routing from u consumes at least $q(\ell - \epsilon)^\alpha$ units of energy. The total battery capacity of i, u, and v equals $(1 + 2 + 1)\ell^\alpha$, as i is an outer edge sensor. This implies $(3 + 2 + 1)q(\ell - \epsilon)^\alpha \le (1 + 2 + 1)\ell^\alpha$. By choosing any ℓ and a small ϵ, the bound $q \le 2/3 + \epsilon_3$ can be obtained for any $\epsilon_3 > 0$.

If the degree of v is at least 2, any routing from u or v to a distant target consumes at least $q(\ell - \epsilon)^\alpha$ units of energy.

Any routing from i consumes at least $2q(\ell - \epsilon)^\alpha$ units of energy; the above simplifying assumption on α implies that the direct path (i, a) is at least as expensive as the multihop path (i, v, a). The total energy consumption equals at least $(1 + 2 + 1)q(\ell - \epsilon)^\alpha$, and the total battery capacity equals $(1 + 3/2 + 1)\ell^\alpha$, as i is an inner edge sensor. The bound $q \leq 7/8 + \epsilon_3$ can be obtained for any $\epsilon_3 > 0$.

Thus, $q \leq 7/8 + \epsilon_3$ in both cases, and relay placement is hard to approximate within $8/7 - \epsilon_2$ for all $\epsilon_2 > 0$. See Table 1 for a summary.

4. HEURISTICS

Despite the lack of efficient approximation algorithms with good approximation ratios, it may still be possible to find a good solution in many practical problem instances by using a heuristic algorithm. However, typical heuristics are of little use in comparing the performance of different kinds of sensor networks: the heuristics provide only a feasible solution (i.e., a lower bound for the optimum); they do not provide an upper bound for the optimum. Thus, seemingly poor performance of a network may be either intrinsic, or it may be caused by the heuristic algorithm finding only suboptimal solutions.

This section presents a (non-polynomial) heuristic method for finding both lower and upper bounds for the solution of the Euclidean relay placement problem. The algorithm maintains an upper bound and a feasible solution. It tightens the upper bound and improves the solution until the ratio of the upper bound and the obtained utility is good enough or a time limit or other termination criterion is met.

In the case of transmission costs defined by power law or a similar monotone function, an optimal solution can be found in the bounding rectangle that contains all sensors and the sink. The upper bound is derived by partitioning the bounding rectangle into a number of rectangular cells. Given a partition, an instance of the finite relay placement problem is constructed.

In the finite problem, the sink and the sensors are the same as in the original Euclidean problem. For each cell, add N possible relays to the same location. The battery capacities of the nodes, the amount of data available at each sensor, the balance parameter, and the reception cost are exactly as specified in the original problem.

Assign a geometrical area to each node. For each sensor node and the sink, this area is the single point of the location of the node. For each relay location, the area is the corresponding rectangular cell. The transmission cost between two nodes is specified as the *lowest possible transmission cost between their respective areas*.

The finite relay placement problem is solved by using any approximate solver. The solver returns a solution x and an upper bound for the utility. The upper bound for the constructed finite problem is also an upper bound for the utility of the original Euclidean problem.

The solution x can be used to construct a solution of the Euclidean relay placement problem by placing each relay at the centre point of the corresponding cell. The utility of this solution can be evaluated by formulating the corresponding balanced data gathering problem. Any LP solver can be used to optimise balanced data gathering [1].

If the tightest upper bound is within a given factor of the best solution, the algorithm terminates. Otherwise, the bounding rectangle is divided into a larger number of cells, and the process is repeated. The following scheme is used here to partition the rectangle. The first partition consists of one cell covering the entire bounding rectangle. At each iteration, split each cell that contains some relay nodes into four new rectangles of equal size. This scheme guarantees convergence while generating only a moderate number of new cells.

The solution returned by the above algorithm is not necessarily a local optimum. Thus, it may be possible to improve the utility of the solution by local search. Here one may use, e.g., line search in a similar way as proposed by Falck *et al.* [1] for their incremental relay placement algorithm.

The algorithm needs to solve the finite problem, either exactly or approximately. One possibility is to formulate the problem as a mixed integer linear program (MIP) and use any MIP solver to find an optimum. The MIP formulation is a relatively straightforward extension to the LP formulation of balanced data gathering [1]. It should be noted that the case of multiple relays at the same location needs to be handled efficiently. This is easy with the MIP formulation: instead of n integral variables with the upper bound of 1, use one variable with the upper bound of n.

The algorithm presented in this paper has been implemented in the C programming language. The source code of the implementation, a test data set, and a number of test results and timings are available[1] under a free software license.

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

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.

5. REFERENCES

[1] E. Falck, P. Floréen, P. Kaski, J. Kohonen, and P. Orponen. Balanced data gathering in energy-constrained sensor networks. In *Proc. 1st International Workshop on Algorithmic Aspects of Wireless Sensor Networks (ALGOSENSORS, Turku, Finland, July 2004)*, volume 3121 of *Lecture Notes in Computer Science*, pages 59–70, Berlin, Germany, 2004. Springer-Verlag.

[2] T. Feder and D. Greene. Optimal algorithms for approximate clustering. In *Proc. 20th Annual ACM Symposium on Theory of Computing (STOC, Chicago, Illinois, USA, May 1988)*, pages 434–444, New York, NY, USA, 1988. ACM Press.

[3] P. Floréen, P. Kaski, J. Kohonen, and P. Orponen. Exact and approximate balanced data gathering in energy-constrained sensor networks. *Theoretical Computer Science*, 344(1):30–46, 2005.

[4] M. R. Garey and D. S. Johnson. The rectilinear Steiner tree problem is NP-complete. *SIAM Journal on Applied Mathematics*, 32(4):826–834, 1977.

[5] B. Krishnamachari. *Networking Wireless Sensors*. Cambridge University Press, Cambridge, UK, 2005.

[6] V. Raghunathan, C. Schurgers, S. Park, and M. B. Srivastava. Energy-aware wireless microsensor networks. *IEEE Signal Processing Magazine*, 19(2):40–50, 2002.

[7] J. Suomela. Computational complexity of relay placement in sensor networks. In *Proc. 32nd Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM, Měřín, Czech Republic, January 2006)*, volume 3831 of *Lecture Notes in Computer Science*, pages 521–529, Berlin, Germany, 2006. Springer-Verlag.

[8] R. Tamassia and I. G. Tollis. Planar grid embedding in linear time. *IEEE Transactions on Circuits and Systems*, 36(9):1230–1234, 1989.

[1] `http://hoslab.cs.helsinki.fi/savane/projects/relays/`