Energy-Efficient Data Collection in WSN: A Sink-Oriented Dynamic Backbone

François Clad, Antoine Gallais and Pascal Mérindol
Image Sciences, Computer Sciences and Remote Sensing Laboratory (UMR CNRS 7005) - University of Strasbourg, France
Email: *{fclad, gallais, merindol}*@unistra.fr

***Abstract*—In wireless sensor networks, energy efficiency is generally achieved by turning off some capabilities from a subset of deployed sensors. The set of active nodes must therefore meet the application requirements (e.g. area coverage, data redundancy) while remaining fully connected to allow further data collection. Here, we focus on the case of a nomad sink entering the network and gathering every monitoring data. At the routing layer, minimizing the number of nodes acting as relays requires to construct a maximum leaf spanning tree (MLST). However, optimizing convergecast communications consists in minimizing the hop distance between the sink and all others nodes, leading so to a shortest path tree rooted at the sink.**

In this paper, we propose a distributed routing protocol that aims at constructing an energy efficient backbone being convergecast efficient at the same time. Our proposal introduces a tradeoff parameter to adjust the compromise "number of relays / routing efficiency" and then constructs a hybrid routing structure based on the combination of variants of the Wu-Li algorithm and a gradient-based routing protocol. For all topologies we simulated, and when tuned for energy saving, our approach outperforms a 2-approximation for constructing a MLST. Furthermore, when tuned for convergecast routing, simulation results show that our solution constructs a routing optimal backbone that involves a small fraction of relays.

I. Introduction

Wireless sensor networks (WSNs) are generally designed for monitoring applications (e.g., disaster prevention or animal tracking, [1]). Such networks do not require any specific infrastructure as sensors themselves act as communication relays. However, a "virtual backbone" is required to enable efficient multi-hop routing protocols. The structure of this backbone should verify several properties to ensure the network efficiency and survivability. In particular, a small-sized backbone allows for saving the global energy consumption. Nodes within the virtual backbone are called *relays* and are expected to form a Connected Dominating Set (CDS) while others are called *leaves*. Thus, minimizing the size of the backbone to form a Minimal CDS (MCDS) turns equivalently the whole graph into a Maximum Leaf Spanning Tree (MLST): the more leaves, the less nodes involved in the routing plan. Subsequently, energy-efficiency can be achieved by finely adapting the communication stack (e.g., at the medium access control layer [2]).

However, only distributed approximations for constructing MLST are practically deployable in WSNs. Indeed, even using a centralized algorithm, the MLST problem is NP-hard [3] and, for such target devices, centralized approximations are not suitable. Furthermore, a convergecast efficient backbone rather consists in minimizing all hop distances between the sink and other sensing nodes. This computation is achieved by constructing a shortest path tree (SPT) rooted at the sink with a protocol such as a gradient [4].

Considering the case of a nomad sink, two communications modes may co-exist. First, when no sink station is around, static sensors communicate for data redundancy or signalization purposes. The emphasis is then only put on energy efficiency. Second, once a nomad sink enters the network, all monitoring data should quickly reach the sink station, thus leading to a sink-oriented backbone. In such a case, the priority is to optimize convergecast routing. Since the two objectives (energy saving and convergecast routing efficiency) may co-exist in WSN deployment, our purpose is to allow static sensors to switch from a pure energy-efficient communication mode to a reliable and efficient data collection routing scheme.

In this paper, we present a distributed protocol able to construct a hybrid routing backbone remaining efficient for both communication modes. Our contribution is parameterizable thanks to a tradeoff variable allowing to force the choice between convergecast routing efficiency and energy saving. We show through simulations the benefits of our proposal. For thousands of topologies we simulated, our solution, when calibrated for energy efficiency, exhibits better performances than a 2-approximation scheme. At the same time, it also allows for constructing an energy efficient backbone for convergecast communications.

The remainder of this paper is organized as follows. Sec. II briefly discusses existing algorithms from which we design and evaluate our solution, as detailed in Sec. III. Finally, based on simulation results, we emphasize in Sec. IV the relevance of our proposal.

II. Background

A. Centralized algorithms for MLST computation

In a WSN where all nodes are distributed over the target area, optimizing energy efficiency requires nodes to self organize into a connected set of active nodes while passive nodes turn into a low-power sleeping mode. To the best of our knowledge, the MLST problem (Sec. II-A) is the most suitable formulation to address this issue. However, as already mentioned, optimal MLST resolution is not tractable and centralized approximations are not practically deployable in WSNs. We use such techniques to position our work regarding (near) optimal results. The MLST problem, or equivalently the

MCDS problem, aims at maximizing the number of leaves while minimizing the number or relays. Therefore, such a tree structure would be the ideal backbone to use for the purpose of energy efficiency. Fujie proposes in [5] an optimal algorithm for constructing a MLST using Linear Programming (LP). We adapt his formulation to implement it in CPLEX [6].Let us introduce some notations. We denote $G = (V, E)$ a connected undirected graph, where V is the set of vertices and E the set of edges. Let $T = (V, E_T)$, with $E_T \subseteq E$, be a spanning tree in G. For a vertex $i \in V$, we denote by $\delta(i)$ the set of all adjacent edges of i in G. For an edge $e \in E$, $x_e = 1$ if $e \in E_T$ and $x_e = 0$ otherwise. Then, the MLST problem can be formulated as follows:

$$\begin{aligned}
\text{Maximize } & \sum_{i\in V} y_i \\
\text{subject to } & \sum_{e\in\delta(i)} x_e + (|\delta(i)|-1)\times y_i \le |\delta(i)| && (\forall\ i \in V) \\
& \sum_{e\in E} x_e = |V|-1 \\
& \sum_{e\in\delta^+(s)} f_e - \sum_{e\in\delta^-(s)} f_e = |V|-1 && (s \in V) \\
& \sum_{e\in\delta^+(i)} f_e - \sum_{e\in\delta^-(i)} f_e = 1 && (\forall\ i \in V\backslash\{s\}) \\
& 0 \le f_{(i,j)} \le (|V|-1)x_e && (\forall\ (i,j) = e \in E) \\
& 0 \le f_{(j,i)} \le (|V|-1)x_e && (\forall\ (i,j) = e \in E) \\
& x, y \in \{0,1\}
\end{aligned}$$

Note that we use a control flow formulation in our LP implementation for efficiency purpose. Using such an implementation, we conduct several experiments to study the scalability of an optimal MLST resolution and thus position distributed heuristics for small WSNs where LP may apply.

However, for large networks, we need to rely on approximation schemes to study the relevance of our proposal. Solis-Oba [7] proposes such a centralized MLST approximation. This method is based on the same concepts as [8]: it consists in constructing a *leafy forest*, made of trees with many leaves. Then, these trees are connected together to form a spanning tree with at least half of the maximum number of leaves. The Solis-Oba algorithm starts by selecting a vertex having a degree greater or equal to 3. An initial tree rooted on it (whose leaves are its neighbors) is thus created. Then, expansion rules are recursively applied on each leaf to extend the tree and preserve a large ratio of leaves. The complexity of this algorithm is in $O(|E|)$ for an approximation ratio of 2. In practice, we implement this algorithm to position our work for large scale WSN simulations.

B. Distributed algorithms for MCDS approximation

The solution whose design is the closest to our approach is a recent contribution of Misra and Mandal [9]. They describe a heuristic for approximating MCDS using collaborative cover. In a first phase, a propagation algorithm is used to construct an independent set within the graph. Then, in a second phase, Steiner [3] nodes are added to this set in order to limit the CDS size. This algorithm achieves an approximation ratio of $(4.8 + \ln 5) \times opt + 1.2$, where *opt* is the size of the MCDS. However, its message complexity of $O(|V| \times \Delta^2)$, where Δ is the maximum vertex degree in the graph, is far too high for low-power environments such as WSNs.

Wu and Li [10] introduce a localized solution for constructing a non bounded approximation of a MCDS. Using solely a local 2-hop knowledge, each node determines its state, *dominating* or *dominated*, and informs its direct neighborhood. A set of pruning rules are used to limit the CDS size while maintaining its connectivity. This algorithm produces a small-sized CDS with a low time complexity of $O(\Delta^2)$. The cost in terms of message exchange is about 1 to 2 messages per node. More recently, several derived solutions have been proposed such as energy-aware [11] and generalized [12] heuristics. In the following, we use the *node degree based rules* presented in [11] as a building block of our proposal.

III. Proposition

A. General presentation

In this paper, we envision the situation where a nomad sink may join the network at any time and desire to quickly gather monitored data. For this purpose, we propose an adaptive solution in order to turn the non sink-oriented routing backbone dedicated to energy saving to a sink-oriented one dedicated to convergecast efficient routing. We aim at creating an opportunistic backbone taking advantage of both the initial MCDS approximation and a sink rooted SPT. This new structure is then designed for convergecast communications while preserving most energy related advantages from the initial one. Indeed, the resulting SPT tends to limit the number of relay nodes required to be fully active.

The tradeoff between both communication schemes can then be defined by a parameter allowing to explore the underlying routing backbone diversity: from the most convergecast efficient model to the most energy efficient one. For that purpose, we introduce a variable, denoted λ and defined in $[1, \infty[$, to make the result tend to one of the spectrum sides. A λ value close to 1 results in an energy efficient SPT generation, while largest values would tend to preserve (and even improve) the preliminary MCDS energy saving advantages.

Practically, we start by applying Wu-Li *node-degree-based* rules detailed in [11] in order to construct a small sized CDS. Then, an elected[1] sensor initiates our enhanced gradient-based protocol with $\lambda \to \infty$. The resulting connected structure will be used for communicating within the network while no sink is declared. Such a backbone may be dedicated for data redundancy purpose and is designed to consume as little energy as possible. Upon entering the network, the nomad sink initiates a gradient construction in order to inform the network of its presence. This gradient message is then forwarded by

[1] In practice, we consider a randomly selected sensor although its selection can impact the resulting CDS size. Such an optimization feature is out of the scope of this paper and let for further works.

all nodes, relays or leaves. Their state is included in gradient messages to be considered during the process. Rather than storing only the best rank received from any node, each node maintains two variables: r_m, giving the best rank received from a marked node, and r_u, depicting the best rank received from an unmarked one. Those marker states are the ones determined by the preliminary Wu-Li MCDS approximation. We detail our gradient-based variant in Sec. III-B.

To summarize, our proposal works as follow:

1) A Wu-Li like algorithm is applied on the initial graph to obtain a preliminary small sized CDS;
2) Our gradient variant using $\lambda \rightarrow \infty$ (and an elected sink) is triggered to prune this preliminary CDS and thus reduce the backbone size;
3) As soon as a nomad sink decides to collect data, it informs the network by triggering our gradient with a λ value reflecting the chosen tradeoff between energy saving and convergecast routing efficiency;
4) Upon leaving the vicinity of the network, the sink notifies its direct neighbors in order to trigger a new gradient with $\lambda \rightarrow \infty$. The network falls back to its previous energy saving configuration.

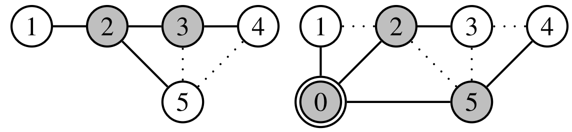
B. Our gradient implementation details

The gradient algorithm [4] is one of the most widespread routing protocols in WSNs. Based on a simple propagation mechanism, it generates a directed acyclic graph rooted at a given sink. Propagated signalization messages carry an integer value giving the number of hops up to the sink, and further referred to as the *rank*. Each node recursively determines its own rank by iterating the lowest received rank.

In our contribution, we use an enhanced gradient algorithm on the top of the Wu-Li localized MCDS approximation. Each node maintains two pairs of initially void variables (r, p), where r gives the rank and p the corresponding identifier of the node delivering this rank. One pair (r_m, p_m) stores information received from the *best marked node*, while the other (r_u, p_u) is used for the *best unmarked node*. The algorithm is initialized by the sink broadcasting the original $GRADIENT(id_{sink},\ 0,\ \lambda)$ message in its vicinity. On receiving a $GRADIENT(source,\ rank,\ \lambda)$ message, a node n increments the variable $rank$, and applies the following procedure:

if $source$ **is** unmarked **and** $rank < r_u$ **then**
  $r_u = rank$; $p_u = source$;
else if $source$ **is** marked **and** $rank < r_m$ **then**
  $r_m = rank$; $p_m = source$;
  if $r_m/r_u > \lambda$ **then**
    send $GRADIENT(n, r_u, \lambda)$
  else
    send $GRADIENT(n, r_m, \lambda)$
  end if
end if

The main idea of our contribution consists in including the parent identifier in gradient messages. Thus each node is aware of its children and knows their identifiers. Childless nodes are then called *leaves* while the others are *relays*. Given that the parent identifier is included in gradient messages, a leaf overhearing its own identifier in a message knows that it should become a relay. To the contrary, a relay not hearing any message with its identifier knows that it has no child and turns into a leaf. This ability is very interesting since it offsets the leaf loss possibly induced by small λ values and noticeably reduces the CDS size for higher values. Experimental results show this feature does not only offset the loss but also improves the leaf/relay ratio in any case ($\forall \lambda \in [1, \infty[$).



(a) Wu-Li preliminary CDS (b) Gradient, node 0 is the sink

Fig. 1. Improve path stretch towards the sink.

C. Benefits of our Gradient on the initial Wu-Li CDS

On the one hand, the Wu-Li CDS may be slightly augmented to reduce hop distances towards the sink when $\lambda \rightarrow 1$. On the other hand, the use of the gradient allows for pruning potential useless cycles in the Wu-Li CDS thanks to the parent identifier added in messages.

Fig. 1 illustrates the main principle of our contribution. The first graph 1a represents a MCDS approximation obtained through the Wu-Li algorithm (this is the first step of our proposal). In this graph, nodes 2 and 3 are marked, while 1, 4 and 5 are not. Then, the nomad sink (node 0) joins the network and initiates a gradient algorithm. Along with gradient messages propagation, the graph is redirected, each node selecting the best ranked marked node, i.e. closest from the sink, within its vicinity to connect with. Consequently, nodes 1, 2 and 5 are directly connected to the sink and node 3 to 2. In this example, we use $\lambda = 1.3$, thus a node may request a neighboring unmarked node to turn into a relay if the best rank received from any marked node is at least 30% greater than the one of the unmarked node. Node 4 meets this requirements since its best marked node provided rank (r_m) is 3 while the unmarked node 5 provides a rank (r_u) of 2. Hence, node 5 turns into a relay (see Fig. 1b). One more relay is required but the hop distance from the sink to node 3 is reduced by 33%. Moreover, node 3 does no longer have any child and becomes a leaf. Eventually, routing paths are optimized for convergecast communications with the nomad sink, using no more relays.

Since the Wu-Li algorithm is a purely localized protocol, it does not prevent from cycles in the resulting CDS. Indeed, each node only knows the links between its direct neighbors and cannot take into account further connections. Two or more *branches* of the graph may therefore be linked in several

(a) The preliminary Wu-Li CDS (b) Gradient, node 0 is the sink

Fig. 2. Prune the CDS using the gradient.

locations, leading to cycles. Hopefully, thanks to its distributed propagation mechanism, the gradient algorithm allows for pruning the Wu-Li based CDS. Unnecessary relays (the ones being no responsible for any leaves) are then turned into leaves. On Fig. 2a, the radio range is equal to the side size of the square. Thus, all nodes remain marked using solely the Wu-Li algorithm. Indeed, each node has two unconnected neighbors and no connected set of direct neighbors covering all of its neighbors. The gradient propagation (from node 0) reveals that marked nodes 3 and 4 do not have any child as no received message included their identifier in the parent field. Consequently, they turn into leaves.

IV. Evaluation

This section deals with the simulations we conducted to evaluate the performances of our proposition. As previously mentioned, we use the Solis-Oba centralized approximation as a near optimal-bound reference to position our proposal. Our simulations were performed using the WSNet [13] event-driven simulator. This software allows to simulate the behavior of large scale WSNs, with message transmissions, energy consumption and mobility. For each communication layer, WSNet provides a modular environment to reproduce realistic conditions using well-known protocols, such as IEEE 802.15.4 and X-MAC [14]. In addition, WSNet comes with a topology generation plug-in that we use to construct random graphs.

For our topology generation, nodes are uniformly spread over a square area, 50 unit long on each side. We generate 100 topologies for each retained density (e.g., the average number of nodes within a communication area). Each node is equipped with an omnidirectional antenna allowing it to communicate directly with all other nodes within a 10 unit radius, using a free-space propagation model. We ensure that all our topologies are connected. Medium access control is handled by the X-MAC protocol. Except for the Solis-Oba centralized algorithm, which is implemented in an external software, all simulated solutions are implemented at network layer. Our implementation of the Solis-Oba algorithm takes as input the same connected graph generated with WSNet and then returns the number of resulting leaves.

A. *CDS size with $\lambda \rightarrow \infty$*

We start with comparing performances (e.g., the backbone size for energy efficiency purposes) of our proposition to those obtained through Wu-Li and gradient algorithms alone and along with the Solis-Oba 2-approximation scheme. Looking at Fig. 3, we can notice that on the commodity hardware

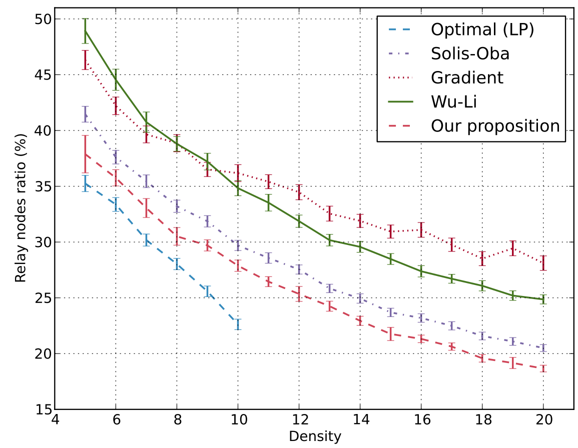


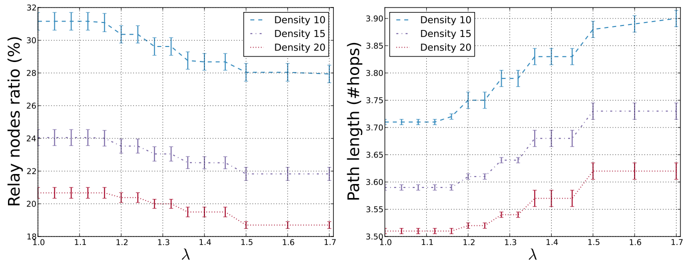
Fig. 3. CDS size comparison.

we use, the scalability of our LP implementation stops for topologies having a density greater than 10. In order to minimize the number of relays in the backbone resulting from our proposal, we use $\lambda = \infty$. With such a calibration, the gradient propagation is mainly used to prune the Wu-Li initial CDS, and, in a lower extent, to re-direct some edges for improving paths length (i.e. number of hops) toward the sink. Results shown in Fig. 3 represent average values over 100 simulations (for each considered density), with a 95% confidence level. Each round of simulation runs on a different randomly-generated topology whereas, within a given round, all algorithms are evaluated on the same topology.

We notice that the gradient propagation produces quite good results (less than 50% relay nodes), considering the fact that it is not intended to produce a small-sized CDS but a sink-rooted SPT. While efficient for sparse topologies, the fraction of relays decreases slower than with others dedicated heuristics. The Wu-Li algorithm produces larger CDSs than the gradient for low-density topologies but the relay ratio quickly decreases as density increases. Indeed, the former may create cycles while the latter does not, which is more apparent with a few number of nodes. We also notice that the difference between the Wu-Li version we retained and the Solis-Oba centralized algorithm also slightly decreases as density increases.

Fig. 3 shows that our hybrid approach clearly outperforms other heuristics, producing, for all simulated topologies, even better results than the centralized 2-approximation. Indeed, for all densities we consider, our algorithm allows to save about $2-3\%$ of nodes in more than with the Solis-Oba algorithm. Moreover, for small density networks where LP apply, we notice that our heuristic produces near optimal CDS although the difference seems to increase with density. Our simulation results show that our contribution is able to produce a small sized backbone that meets our energy efficient requirement.

Wu-Li generated ones, on top of being suited for convergecast communications. The low number of extra relays nodes is so limited that $\lambda = 1$ would perform the best calibration to speed up the data collection when a sink enters the network.



(a) CDS size (b) Path stretch

Fig. 4. Tradeoff adjustment using λ.

B. CDS size and path stretch with $\lambda \rightarrow 1$

As previously stated, when no sink is declared and for all generated topologies, our heuristic generates better results than a 2-approximation. However, this result does not reflect the convergecast routing efficiency, i.e., the quality in terms of path length. With $\lambda \rightarrow \infty$, we just aim at approximating a MCDS and prevent any leaf from turning into a relay, thus hampering path length reduction. Here, we aim at reaching a tradeoff between a small-sized backbone and an efficient routing scheme to quickly gather monitored data to the sink (for both low energy consumption and communication efficiency).

We therefore evaluate the impact of λ on the fraction of relays and on the path lengths. Fig. 4a gives the fraction of relays according to λ while Fig. 4b provides the average distance to the sink according to λ. For the sake of clarity, on both figures, we decide to take interest in three given densities: 10, 15 and 20. Note that sinks are randomly placed for those simulations (a single sensor is randomly selected to act as the closest one).

First of all, note that for $\lambda = 1$, we obtain better results, in terms of CDS size, than the Wu-Li algorithm alone. When targeting data collection efficiency, the number of new involved relays is balanced by our gradient based graph pruning. Indeed, our hybrid method calibrated with $\lambda = 1$ is not identical to a sole gradient algorithm: nodes select their parent according to their impact on the CDS size. When there is a rank equality between a relay and a leaf node, our technique permits to select the relay first for tie-breaking. The initial Wu-Li algorithm allows to assign a priority to each node, thus reducing the global number of relays in the final backbone.

Fig. 4a shows that the relay ratio remains stable for $\lambda < 1.05$ before quickly decreasing and stabilizing again when λ reaches 1.5. For a density of 20 and $\lambda = 1$, we notice that the CDS size increases only about 3% compared to $\lambda \rightarrow \infty$. Focusing on the path stretch improvement in Fig. 4b, we observe that both stabilization thresholds visible on Fig. 4a are still present. We notice that a relaxed λ value allows to decrease the distance by 3% to 5%. Note that this relative gain of $\approx$ 4% is not uniformly distributed over path lengths: the longer the original path, the higher is the gain.

Our proposition allows to reduce the average path length by more than 4% using only up to 3% extra relays (compared with $\lambda \rightarrow \infty$). Resulting backbones are always smaller than Wu-Li generated ones, on top of being suited for convergecast communications. The low number of extra relays nodes is so limited that $\lambda = 1$ would perform the best calibration to speed up the data collection when a sink enters the network.

V. Conclusion

In this paper, we first summarize main centralized solutions to construct optimal or near optimal MLST: we implement them to evaluate our proposition. We then focus our attention on practical distributed heuristics producing small sized routing backbone while satisfying WSN requirements (low time complexity and communication cost). Our proposition consists in using an enhanced gradient algorithm on top of a variant of the Wu-Li algorithm. It results in a hybrid backbone achieving the desired compromise between energy saving and optimal convergecast routing. Our version of the gradient protocol is able to prune the Wu-Li CDS to reduce the fraction of relays while, at the same time, reducing path length for convergecast communications. For every simulated topology, our solution systematically outperforms a 2-approximation algorithm when calibrated for energy efficiency. We also show that our hybrid approach achieves an interesting tradeoff between path optimality and fraction of relays node when oriented for data collection. Our future works will focus on implementing it on our large scale SensLAB testbed [15].

References

[1] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer Networks*, vol. 52, no. 12, 2008.

[2] R. Kuntz, A. Gallais, and T. Noel, "From versatility to auto-adaptation of the medium access control in wireless sensor networks," *Journal of Parallel and Distributed Computing*, vol. 71, no. 9, 2011.

[3] M. Garey and D. Johnson, *Computers and intractability: a guide to the theory of NP-completeness*, Freeman ed., 1979.

[4] H. Karl and A. Willig, *Protocols and Architectures for Wireless Sensor Networks*, Wiley-Interscience ed., 2007.

[5] T. Fujie, "An exact algorithm for the maximum leaf spanning tree problem," *Computers & Operations Research*, vol. 30, no. 13, 2003.

[6] IBM ILOG, "CPLEX Optimizer, high-performance mathematical programming solver for linear programming, mixed integer programming, and quadratic programming," http://www.ibm.com.

[7] R. Solis-Oba, "2-Approximation algorithm for finding a spanning tree with maximum number of leaves," in *Algorithms - ESA' 98*, Springer Berlin / Heidelberg ed., vol. 29, no. 1, 1998.

[8] H. Lu and R. Ravi, "Approximating maximum leaf spanning trees in almost linear time," *Journal of Algorithms*, vol. 29, no. 1, 1998.

[9] R. Misra and C. Mandal, "Minimum connected dominating set using a collaborative cover heuristic for ad hoc sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 3, 2010.

[10] J. Wu and H. Li, "A Dominating-Set-Based routing scheme in ad hoc wireless networks," *Telecommunication Systems*, vol. 18, no. 1, 2001.

[11] J. Wu, F. Dai, M. Gao, and I. Stojmenovic, "On calculating power-aware connected dominating sets for efficient routing in ad hoc wireless networks," *Journal of communications and networks*, vol. 4, no. 1, 2002.

[12] F. Dai and J. Wu, "An extended localized algorithm for connected dominating set formation in ad hoc wireless networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 15, no. 10, 2004.

[13] G. Chelius and al., "WSNet, an event-driven simulator for large scale WSNs," http://wsnet.gforge.inria.fr.

[14] M. Buettner, G. V. Yee, E. Anderson, and R. Han, "X-MAC: a short preamble MAC protocol for duty-cycled wireless sensor networks," in Proc. of *4th ACM International Conference on Embedded Networked Sensor Systems*, Oct. 2006.

[15] Senslab consortium, "SensLAB, a very large scale open wireless sensor network testbed," http://www.senslab.info.