

Multiple Sink Network Design Problem in Large Scale Wireless Sensor Networks

E. Ilker Oyman and Cem Ersoy

Computer Engineering Department, Bogazici University, Istanbul, Turkey
{oymani, ersoy}@boun.edu.tr

Abstract— The battery resource of the sensor nodes should be managed efficiently, in order to prolong network lifetime in wireless sensor networks. Moreover, in large-scale networks with a large number of sensor nodes, multiple sink nodes should be deployed, not only to increase the manageability of the network, but also to reduce the energy dissipation at each node. In this paper, we focus on the multiple sink location problems in large-scale wireless sensor networks. Different problems depending on the design criteria are presented. We consider locating sink nodes to the sensor environment, where we are given a time constraint that states the minimum required operational time for the sensor network. We use simulation techniques to evaluate the quality of our solution.

Keywords—wireless sensor networks; power efficiency; multiple sink.

I. INTRODUCTION

WIRELESS sensor nodes are combining the wireless communication infrastructure with the sensing technology. Instead of transmitting the perceived data to the control center through wired links, ad hoc communication methods are utilized, and the data packets are transmitted using multi-hop connections.

The efficiency of the sensor network investment is directly related with the length of the reliable monitoring duration of the field. The better energy control mechanisms are used in the sensor nodes' firmware and in the network management techniques, the longer the network will be serving their investors. Therefore, the limited battery resource of the sensors should be handled efficiently.

In some applications, several thousands of sensor nodes might be deployed over the monitored region. For example, in agricultural scenarios, in environmental monitoring or in border surveillance applications, such large-scale sensor networks would be necessary. Moreover, the diameter of the region might easily be several kilometers. In this case, scalability of the network is a very important design issue. In order to obtain a scalable network, the sensor nodes should be divided into clusters. The nodes within a cluster will then be connected to the sink nodes dedicated for that cluster. Fig. 1

shows such a sensor network with several nodes and three clusters with three sink nodes.

During the design phase of a large-scale sensor network, the designer should decide on the number of clusters, and more important than that, the optimum locations of the sink nodes. We call this problem as the “multiple sink sensor network design problem” and try to provide some solutions.

The problem of finding the number and the location of the sink nodes resembles some classical problems like plant location problem [1], warehouse location problem [2], and the concentrator location problem (CLP), which have been received a great interest in the literature [3], [4]. There are, however, several differences between these problems and the multiple sink location problem. First, the alternatives of the sink locations are not enumerated, i.e., the sinks can be placed everywhere in the environment. Second, the communication between the sensor nodes and the sink nodes are performed using multi-hop links rather than direct connections. Third, the capacity of the sink nodes is only related with the data processing power of the sink node, rather than the amount of available ports in the concentrator.

The organization of the paper is as follows: In Section 2, we state the “multiple sink sensor network design problem” together with their design issues that are important for large scale wireless sensor networks. In Section 3, a solution technique this problem is presented. Section 4 shows computational experiments with a demonstrative example for the sink location problem. We conclude the paper in Section 5.

II. SINK LOCATION PROBLEMS

The amount and the exact locations of the sink nodes directly affects the lifetime of the sensor network. Therefore,

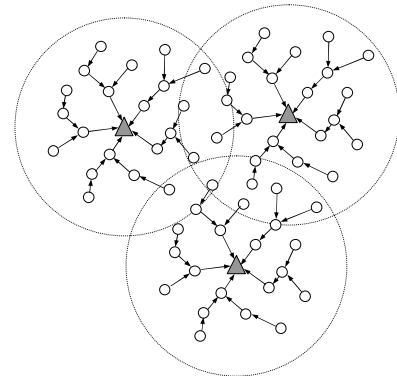


Fig. 1. A large-scale sensor network with three clusters

This work was supported in part by the State Planning Organization of Turkey (DPT) under the grant number DPT98K120890, and by Opnet Technologies, Inc. through “Teaching with OPNET” program.

for an economically feasible investment, the designer should focus on correct placement of the sink nodes. Depending on the design criteria, there might be several approaches for this problem.

A. Find the Best Sink Locations (BSL)

In many real world deployment scenarios, the designer will have a predefined budget granted for the investment. Therefore, the number of sink nodes is known prior to the deployment phase. Since we know the number of sink nodes, which represents the number of clusters in the network, the only problem remains is the efficient clustering of these sensor nodes. We call this problem as finding the “Best Sink Locations” problem (BSL).

There are many good clustering algorithms in the literature, which are mainly classified as hierarchical and non-hierarchical methods [5]. The non-hierarchical methods are usually referred to as k-means clustering [6]. Another generic method is the self-organizing maps, which is a general-purpose unsupervised learning algorithm [7]. None of these methods provide the optimum number of clusters that should be formed. The number of clusters should be given as a decision parameter to the algorithms.

Exact location of the sink nodes are easily found when the clustering algorithm completes. Whenever the Euclidean distance is used as the clustering metric, then the center of mass of the nodes within a cluster would give the location of the sink nodes. Depending on the priorities of the routing algorithm, power aware distance metrics could also be used.

B. Minimize the Number of Sinks for a Predefined Minimum Operation Period (MSPOP)

In some applications, the investor might request the sensor network be operational for a predefined duration. For example, in agricultural applications, the field must be monitored until the harvest. Therefore, the sensor network should be reliable until the crops grow up, and are reaped from the field. The farmer is definitely going to deploy a new sensor network in the next season. We call this problem as “Minimization of the number of Sink nodes for a Predefined minimum Operation Period” (MSPOP).

In order to solve this problem, we have to calculate the sensor network lifetime for any number of sink nodes. Then, only the solution will be selected, where the network lifetime exceeds the predefined limiting constraint, with the minimum number of sink nodes.

The major issue in this problem is to select the correct number of sink nodes. The brute force technique is to start with only one sink node, as stated in [5]. While incrementing the number of sink nodes by one, the network lifetime is evaluated. The search will stop, whenever the desired lifetime is reached.

C. Find the Minimum Number of Sinks while Maximizing the Network Life (MSMNL)

We may also try to extend the network lifetime as much as possible with the most economical investment. We call this problem as “Minimization of the number of Sink nodes while Maximizing the Network Lifetime” (MSMNL). When we do not have a prior knowledge on the number of sink nodes nor the lifetime constraint, then we bring up a combinatorial optimization problem.

In this case, the objective should combine the two alternatives within one function, where the budget reserved for the sink nodes should somehow be connected with the lifetime of the sensor nodes. The initial investment for the sensor nodes should be utilized for the longest period. Here, the “cost per unit time” metric could be used. In order to reach the required timeframe, we may need to perform some redeployment within the network. We should add the cost of the redeployment to the initial investment cost, which includes both the sensor and the sink nodes. The near-optimal solution for this problem could be found using some heuristic techniques, like simulated annealing or genetic algorithms.

III. A SOLUTION TECHNIQUE FOR THE MSPOP PROBLEM

After discussing the issues related to the multiple sink sensor network design and listing alternative problems, in this section, we will consider particularly the MSPOP problem and propose a solution technique.

This problem is by nature an off-line problem, which should be solved by the system designer at a central location. Therefore, the location information of the sensor nodes should be collected from the field, before the solution phase. The system design algorithm is given in Fig. 2. In the following sections, each step of this algorithm is explained in detail.

A. Deployment of the Sensor Nodes

Depending on the underlying application, there may be several alternatives on sensor network deployment. In an agricultural application, they might be scattered by the farmer, in a nearly uniform manner. In an environmental application like forest fire detection, they might be dropped from an aircraft. In in-house applications, they might be installed by the construction workers by hand. Nevertheless, the clustering algorithm that is used at Step 5.ii in Fig. 2 should be able to deal with these deployment scenarios.

-
1. Deploy the sensor nodes
 2. Wait until the sensor nodes find their location information
 3. Collect the location information from the field
 4. $k = 0$
 5. Repeat
 - i. $k = k + 1$
 - ii. Find the best location for k sink nodes
 - iii. Estimate the network lifetime
 6. Until required network lifetime is reached
 7. Output the sink locations, and the estimated lifetime
-

Fig. 2. System design algorithm

B. Finding Location Information

In order to calculate sink locations, we must know the location of each individual sensor node. Location information can easily be derived, using central or distributed methods (see [8],[9]). If we use a central approach, we can continue our algorithm with Step 3 in Fig. 2, where the central agent can perform an off-line location estimation.

C. Collecting the Location Information from the Field

After the deployment of the sensor nodes, we can use a mobile terminal to collect the location data, which traverses the field. If we do not have such a special purpose terminal, then we have to install a sink node temporarily. We can take this sink node back during the final deployment phase. When this is not possible due to physical limitations of the application, then we can choose to use this node as a fixed sink node in the final deployment, and consider this node in the clustering algorithm, without moving it.

D. Finding the Best Location for K Sink Nodes

If the number of sink nodes is known, then we have the problem defined in Section II. The clustering algorithm is responsible to locate the sink nodes optimally. Depending on the deployment distribution, we can choose different clustering algorithms at this step. In our implementation, we have used the well known k-means clustering algorithm [6].

E. Estimating the Network Lifetime

In order to estimate the network lifetime, we have to define a metric related to the reliability of the data retrieved from the environment. In our implementation, we use the ratio in (1) as the reliability measure, which is a monotonically increasing function. When a predefined threshold value is reached we say that the network is not reliable anymore.

$$\rho(t) = \frac{\text{number of unreachable nodes at time } t}{\text{total number of nodes}} \quad (1)$$

We use a simulator to find out when the predefined reliability threshold is exceeded for the estimation of the lifetime. This simulator should consider the packet generation rate, which is an application criterion, sensor node's hardware specifications, like the battery capacity, communication data rate, overhead energy, transmitter power requirements, and environmental characteristics, like the path loss exponent.

IV. COMPUTATIONAL EXPERIMENTS

In this section, we try to give two examples of multiple sink sensor network design problems. First, we will provide a demonstrative example for the BSL problem on a sample sensor network with three sink nodes. After that, we will consider the MSPOP problem and show the application of the solution technique described in Section III with several sensor network deployments.

A. Simulation Setup

In our simulations, we only use sensor nodes having power adjustable transmitter circuitry. Total energy dissipation e_j^T of the sensor node j is modeled as a function of distance d as follows.

$$e_j^T(d) = \kappa d^\alpha + \tau \quad (2)$$

where $\kappa, \tau \in \mathfrak{R}$ are real numbers. τ is the *overhead energy*, representing the sum of the receiver, sensing and computation energy which is a constant value with varying distance d .

The sensor nodes have a power level adjustment capability on a continuous scale between the minimum and the maximum transmission power values. In simulations, however, we have used 20 discrete power levels instead, to approximate continuity. The simulations are performed on the test bed developed using Opnet Modeler 9.1 [10] to estimate the effective network lifetime.

The sample sensor network consists of 200 nodes which are distributed uniformly over a planar square region with 200 m x 200 m dimensions. The sensors are assumed to use 800 mW transmission power for a 200 m radio range in open air ($\alpha = 2$). These values are chosen, as they are very close to the Berkeley/Crossbow Mica Motes' specifications [11]. However, we have scaled the radio range for our simulation environment where we used a constant path loss exponent value with $\alpha = 3$. For the effect of different path loss exponent values on energy dissipation, the reader may refer to [12]. The energy model in (2) is used to calculate the energy spent at each sensor node for one packet transmission. Here, the overhead energy τ is chosen to be 20 mJ per packet, where 400 mW receiver power is assumed. Both sensing and computation energies are neglected, since they do not affect our design decisions. The data rate of the communication channel is chosen to be 20 kbps, and a fixed packet size of 1024 bits is used. The initial battery capacity of the sensors is chosen to be 200 J. In [13], it is given that for an alkaline-manganese dioxide battery, the typical volumetric energy density is 428 Watt hour per liter. In other words, a battery of size one cubic centimeter would have the capacity 1540 J. However, we have chosen the smaller value to shorten the simulation time. The behavior of the simulations will not change, since the battery capacity only causes the results to appear earlier. As a result, we have also chosen a short simulation time of 60 days.

The lifetime is defined to be related with the number of unreachable sensors in the network. We consider the readings of the network to be unreliable anymore, whenever the ratio of the unreachable sensors exceeds this threshold value, with $\rho = 0.25$. The sensors are assumed to perform independent readings, and therefore independent packet generations. The packet generation process is assumed to be a Poisson process with rate $\lambda = 1$ packets per hour, where a continuous monitoring application is assumed. Nevertheless, here a

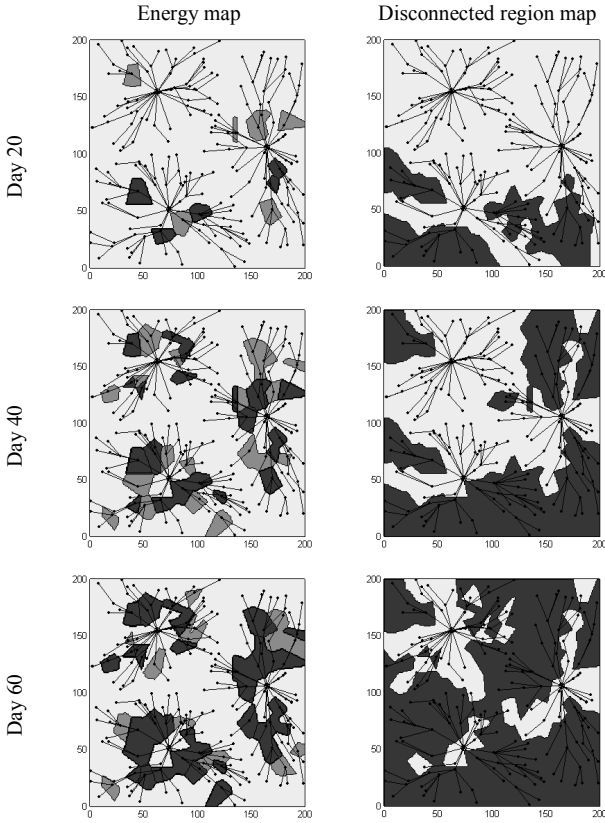


Fig. 4. Energy and disconnected region maps, until the 60th day

periodic process could also be chosen where the sensors are polled with a predefined frequency.

B. Demonstrative Example for the BSL Problem

In this problem, we are given a sensor network. We try to locate three sink nodes accordingly to reach the maximum operation time. The locations of the sink nodes are found using the well known k-means clustering algorithm [6].

After the deployment phase, we tried to estimate the network lifetime. For this, we constructed the routing tree, where the minimum energy tree approach is used, with the energy metric given in (2). Then, we monitored the energy map and the disconnected region map of the network. The energy map indicates the regions, where an energy shortage occurs, using iso-energy contours. The darker the contours are, the less energy resource is present in that region. After a period, some nodes start to become exhausted, and therefore all the nodes in the branch served by this node will be disconnected. The disconnected region map shows these disconnected nodes due to energy failures. In Fig. 3, we show these energy and disconnected region maps, where we have taken snapshots of the sensor network once every 20 days.

We observe here, that the disconnected region increases, as time goes by. This behavior is expected, since the energy reserve of the sensors decrease during operation. Energy shortage occurs at nodes, who serve a large branch as a relay node. None of the leaf nodes encounters an energy problem, however, some of them cannot reach the sink nodes as they

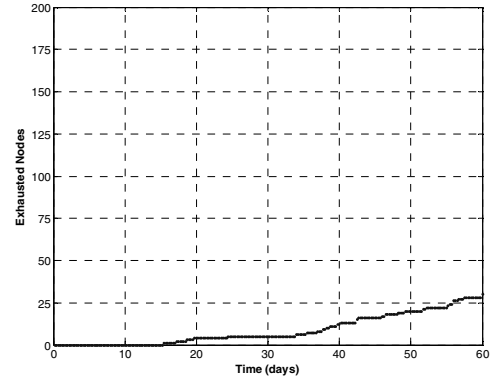


Fig. 5. Exhausted nodes versus time

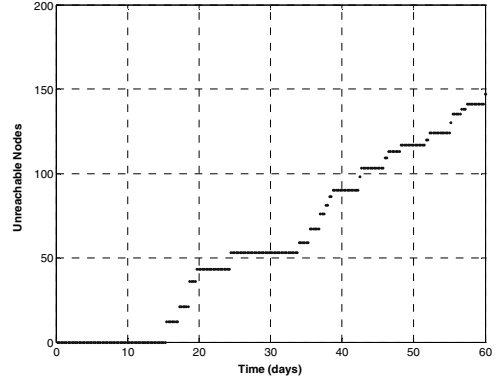


Fig. 3. Unreachable nodes versus time

become disconnected because of failures in the relay nodes. We observe early failures at sensors that are close to the sink nodes, since they serve a larger branch set. On the other hand, not every sensor node that is close to sink nodes fail, since these nodes may have a very small number of sensor nodes to serve. Therefore, we can expect that reconstructing the minimum energy tree after energy failures could prolong the network lifetime.

In Fig. 4, we show the increase on the number of exhausted nodes. We observe the first failure at the 15th day. The failed node has 12 nodes in its branch set. Therefore, the first failure causes 11 other functional nodes to become disconnected. The number of unreachable nodes can be seen in Fig. 5. Using this figure, we can state that the network produces reliable readings only for 24 days, where we pass the threshold for the number of disconnected nodes, with $\rho = 0.25$.

C. Application of the Solution Technique to the MSPOP Problem

In this section, we perform simulations on multiple, uniformly deployed sensor networks, and try show the application of the solution technique for the MSPOP problem. We start with one sink node, and increase the number of sink nodes by one at every step, where the location of the sink nodes are found using the k-means clustering algorithm [6]. After each deployment alternative, we tried to estimate the network lifetime. We have considered up to six sink nodes in this simulation set.

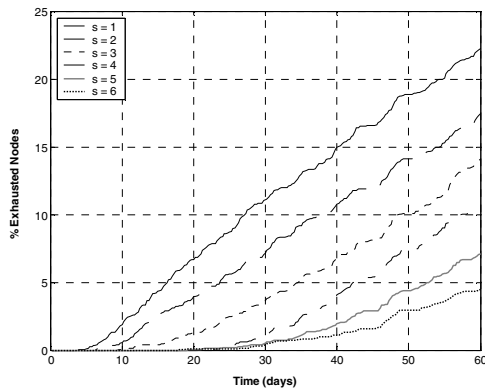


Fig. 6. Percentage of exhausted nodes versus time, with different number of sinks

In Fig. 6, we show the percentage of exhausted nodes, where the number of sink nodes is varying from one up to six. The cluster size decreases with increasing number of sink nodes. Therefore, the paths from each sensor node to the sink nodes will be shorter in the resulting minimum energy tree. As a result, energy dissipation due to packet relays decreases, therefore the percentage of exhausted nodes decrease.

In Fig. 7, the percentage of unreachable nodes is shown. When the number of sink nodes is small, then the percentage increases very rapidly within the few days. Then, only the nodes that are close to the sink nodes survive. Moreover, since these nodes have a relatively less relaying load, their batteries are exhausting rather slow.

As the solution of our problem with $\rho = 0.25$, we have to be given a minimum desired network lifetime. If this period is required to be one month, then it is clear that we have to utilize four sink nodes. When this period is required to be at least two months, the number of required sink nodes increases to six.

V. CONCLUSION

In order to maximize the lifetime of a sensor network, energy resources of each individual sensor node must be consumed effectively. In large-scale sensor networks, the network must be divided into smaller sub-networks, not only to increase manageability of the network, but also to increase the network lifetime. We have introduced the multiple sink network design problem, where the best places for the sink nodes should be calculated depending on several different design criteria.

We have demonstrated a sample sink location case, where the number of sink nodes was known before the deployment phase. We have implemented the solution for this BSL problem, and presented the corresponding energy and disconnected region maps on a sample sensor network for different snapshots in time. We have observed how the disconnected region increases with time. We have encountered with failures at the sensor nodes that are close to the sink nodes, because these nodes have served a larger branch set. However, not every sensor node that is close to sink nodes fail early, since these nodes might coincidentally have a very small

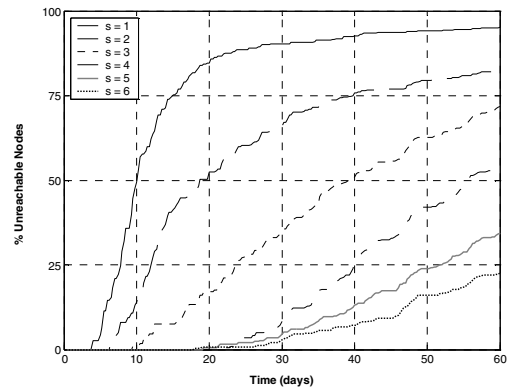


Fig. 7. Percentage of unreachable nodes versus time, with different number of sinks

number of sensor nodes to serve. So we can conclude that reconstructing the minimum energy tree after energy failures occur could prolong the network lifetime. We have also proposed a solution technique for the MSPOP problem, and simulated its implementation on random networks. We have analyzed the effect of adding new sinks to the network lifetime. We have presented methods to deploy economically feasible amount of sink nodes while prolonging the network lifetime.

REFERENCES

- [1] A. Alcouffe, and G. Muratet, "Optimal Location of Plants," *Management Science*, Vol. 23, No. 3, pp. 267-274, Nov. 1976.
- [2] U. Akinc, and B. M. Khumawala, "An Efficient Branch and Bound Algorithm for the Capacitated Warehouse Location Problem," *Management Science*, Vol. 23, No. 6, pp. 585-594, Feb. 1977.
- [3] R. R. Boorstyn, and H. Frank, "Large-Scale Network Topological Optimization," *IEEE Transactions on Communications*, Vol. com-25, No. 1, pp. 29-47, January, 1977.
- [4] B. Gavish, "Topological Design of Computer Communication Networks-The Overall Design Problem," *European Journal of Operations Research*, Vol. 58, No. 2, April, 1992.
- [5] J. J. F. Hair, R. E. Anderson, R. L. Tatham and W. C. Black, *Multivariate Data Analysis with Readings*, Prentice-Hall, Inc., 1995.
- [6] J. A. Hartigan, *Clustering Algorithms*, Wiley, New York, 1975.
- [7] T. Kohonen, *Self-Organizing Maps*, Springer, 1995.
- [8] A. Savvides, H. Park and M. B. Srivastava, "The Bits and Flops of the N-hop Multilateration Primitive for Node Localization Problems," *Proceedings of the First ACM International Workshop on Wireless Sensor Networks and Applications (ACM WSNA 2002)*, pp. 112-121, Atlanta, Sept. 28, 2002.
- [9] A. Nasipuri, and K. Li, "A Directionality based Location Discovery Scheme for Wireless Sensor Networks," *Proceedings of the First ACM International Workshop on Wireless Sensor Networks and Applications (ACM WSNA 2002)*, pp. 105-111, Atlanta, Sept. 28, 2002.
- [10] Opnet Technologies, Inc., *Opnet Modeler Home Page*, <http://www.opnet.com/products/modeler/home.html>, 2003.
- [11] Crossbow Technology, Inc., *Wireless Sensor Networks, TinyOS, Berkeley Motes from Crossbow*, http://www.xbow.com/Products/Wireless_Sensor_Networks.htm, 2003.
- [12] E. I. Oyman, *Multiple Sink Location Problem and Energy Efficiency in Large Scale Wireless Sensor Networks*, Ph.D. Thesis, Bogazici University, 2004.
- [13] Duracell, *Alkaline-Manganese Dioxide Technical Bulletin*, <http://www.duracell.com/oem/Pdf/others/ATB-full.pdf>, 2003.