Cooperative Data Collection in Ad Hoc Networks*

Liron Levin Michael Segal Hanan Shpungin

Abstract

This paper studies the problem of data gathering in multi-hop wireless ad hoc networks. In this scenario, a set of wireless devices constantly sample their surroundings and initiate report messages addressed to the base station. The messages are forwarded in a multi-hop fashion, where the wireless devices act both as senders and relays. We consider data gathering without aggregation, i.e. the nodes are required to forward *all* the messages initiated by other nodes (in addition to their own) to the base station. This is in contrast to the well studied problem of data gathering with aggregation, which is significantly simpler.

As some nodes experience a larger load of forward requests, these nodes will have their battery charges depleted much faster than the other nodes – which can rapidly break the connectivity of the network. We focus on maximizing the network lifetime through efficient balancing of the consumed transmission energy. We show that the problem is NP-hard for two network types and develop various approximation schemes. Our results are validated through extensive simulations.

1 Introduction

Wireless ad hoc networks have found their way into almost every advanced technology in the market; among those are mobile communication, radio broadcasting, and sensor monitoring. The temporary physical topology of the network is determined by three factors: the distribution of the wireless transceivers, their transmission ranges, and obstacles in the deployment area (e.g. buildings). These three factors constitute the directed communication graph where the nodes correspond to the transceivers and the edges correspond to the communication links. A communication link from node u to node v is established if the Euclidean distance between u and v, $d(u, v)$, is less than the transmission range of u and there are no obstacles which might interfere with the transmission.

In this paper we explore the well-known *data gathering* network scheme, i.e. each node collects information from its surrounding area and then propagates it in a multi-hop fashion, using other nodes as relays, to some base station, also referred to as the *root node*. Many important applications benefit from this data gathering scheme, such as habitat monitoring [28], security applications [3], and civil structure monitoring [6]. The information each node collects is encoded into messages, which are then propagated by using a *data gathering*

*A preliminary version of this paper has appeared in WiOPT'10 [23]. The work on this paper has been partially supported by US Air Force European Office of Aerospace Research and Development, grant #FA8655-09-1-3016, Deutsche Telecom, European project FLAVIA and Israeli Ministry of Industry, Trade and Labor (consortium CORNET). Liron Levin and Michael Segal are with the Dept. of Communication Systems Engineering, Ben-Gurion University of the Negev, Israel. Hanan Shpungin is with the Dept. of Computer Science, University of Calgary, Canada.

tree ([12, 38, 37]). The data gathering tree is a subgraph of the directed communication graph, such that there is a unique directed path from every transceiver to the base station.

The general data gathering problem is to find an efficient data gathering tree, where efficiency is measured according to the desired optimization objective. One of the most common optimization objectives studied in the context of data gathering is *energy efficiency* ([38, 2, 27, 24, 40, 26]). As wireless nodes are usually equipped with non-replenishable batteries, low energy consumption is one of the most important challenges faced by a network designer. In this work we focus on *network lifetime* maximization, which is the time that the first node runs out of its battery charge, and thus the data gathering tree can no longer be used ([20, 29]).

The problem of data gathering can be divided into two major paradigms. Data gathering *with aggregation* ([17, 34]) allows each node to accumulate the messages of its descendants and then pass only one fixed-size message to its ancestor in the data gathering tree. The second paradigm, is data gathering *without aggregation* ([5, 24]) which requires that *all* messages initiated by the wireless devices eventually will reach the base station.

Surprisingly, data gathering with aggregation is considerably easier to solve under the energy efficiency objective ([32, 31, 17]) than its unaggregated counterpart ([2, 35, 27, 24]). One of the possible reasons for the difference may lie in the fact that once data is allowed to be aggregated, the efficiency of the solution is dictated by the chosen communication links since every node transmits only once. Thus, the problem can be usually reduced to a simpler problem in graph theory ([32, 31]) and solved by using some of the existing methods. Once data aggregation is not allowed, one has to take the amount of data into account as well, since nodes are required to make multiple transmissions over the same communication links.

Practically, in data gathering without aggregation, the load on the nodes that are close to the base station is relatively high; in addition, the time it takes for all messages to reach the base station can be quite long (since the packets can collide). One possible solution is to use a multi-code multi-packet transmission schema (such as the one presented in [41]), which decreases the time until all the messages are received. Another solution is to use a fast data transmission schema which keep transmission energy relatively low (see [15]). In addition, the network can be partitioned to several clusters, with one leader chosen per cluster, such that the data is aggregated to the network leaders and then a separate crawler collects the data from them (such model has been suggested in [7]). For example, root's children may serve as the leaders of their clusters (their respective subtrees).

Our main contribution in this paper is the study of the data gathering problem with an optimization objective of network lifetime maximization. We consider two network models: **homogeneous networks with obstacles** and **heterogeneous networks without obstacles**. In the former case, all nodes share the same transmission range[1], and there are obstacles in the deployment area which prevent some nodes to communicate even if they are within the transmission range of each other (urban area). In the latter case, the nodes may have different transmission ranges and there are no obstacles (open field). The communication graph is determined by the distribution of nodes and the transmission range of every node.

The rest of the paper is organized as follows. In Section 2 we define the system settings and present a formal problem formulation. Then, in Section 3 we discuss some of the previous work and state our contribution. Sections 4 and 5 address the homogeneous and heterogeneous models, respectively. Section demonstrates the numerical results obtained from simulations. Finally, we conclude and propose possible future research directions in Section 7.

2 System settings and problem formulation

In this section, we define our wireless network model and present the data gathering problem under two network scenarios considered in this paper.

2.1 Wireless network model

The network consists of n wireless devices (nodes), V, positioned in a two-dimensional Euclidean plane. In addition there is a base station, r, also referred to as the root node. The data gathering process is executed in discrete rounds. In every round each node u, has $q(u)$ messages to send to the root node r. The messages are propagated towards the root node in a multi-hop fashion by using a *convergecast tree* $T = (V, E_T)$ (also referred to as the *data gathering tree*), where all the edges point towards the root r, i.e. T is a reversed arborescence rooted at r. The same tree T is used for all the rounds.

Let $V(u)$ represent the descendants of u in T. We define $T(u)$ to be the subtree rooted at u such that all the nodes in $T(u)$ pass their messages to u. Eventually, in every round, every $u \in V$ forwards $cn(T,u)$, $cn(T,u) = \sum_{v \in V(u)} q(v)$, messages to its ancestor, $\pi(u)$, in T. We refer to $cn(T,u)$ as the *congestion* of u. Note that according to this model, the inner nodes forward all the messages which originate in their respective subtrees.

Let $\omega(u,v)$ be the amount of energy (also referred to as the *cost*) required to transmit a single message from u to v. We define the cost of a node $u \in V$, $C(T,u)$, as the total energy consumed by u in a single round as a result of transmitting $cn(T,u)$ messages, i.e. $C(T,u) = cn(T,u) \cdot \omega(u, \pi(u))$.

Each node $u \in V$ has an initial battery charge of $b(u)$, which is reduced by $\omega(u, \pi(u)$ after a transmission of a single message. Recall that the same tree T is used in all rounds and the lifetime of a node, $l(u)$ is defined as the number of *complete* rounds in which it can participate, i.e. $l(u) = \lfloor b(u)/C(T,u) \rfloor$ [20, 29]. The network lifetime, $l(T)$ is defined as the first round in which a node cannot complete all its transmissions,

[1] The homogeneous model is better known as the UDG model [2].

i.e. $l(T) = \min_{u \in V} l(u)$. We assume that all initial battery charges are equal.

2.2 Problem formulation

The transmission capabilities of a wireless network are derived from the relative disposition of the wireless nodes and various obstacles in the deployment area. Let $G = (V, E)$ be the transmission capabilities graph (also referred to as the *input graph*), i.e. if $(u, v) \in E$ then node u is able to transmit a message to node v (recall that this transmission costs $\omega(u, v)$). The general maximum network lifetime data gathering problem is formulated as follows:

Problem 1 (Maximum Lifetime Data Gathering Problem)**.** *Given a set of wireless nodes V and a base station r in the Euclidean plane and a transmission capabilities graph G, find a convergecast tree T rooted at r such that $l(T)$ is maximized.*

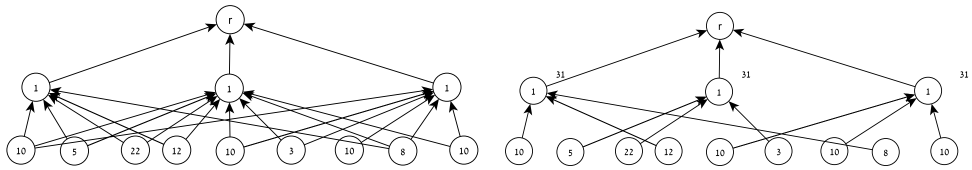
Clearly, in order to have a feasible solution for the maximum lifetime data gathering problem (abbreviated by the *data gathering problem*), there must be a directed path from u to r in G for every $u \in V$. In what follows we define the two network scenarios considered in this paper. The scenarios differ by their definition of the cost function ω and the graph G.

2.2.1 Homogeneous network with obstacles

In this scenario, all nodes share the same transmission range, R and are able to communicate only with nodes which are within this transmission range. Let $G_1 = (V, E_1)$ be the transmission capabilities graph, such that $(u, v) \in E_1$ iff $d(u, v) \leq R$ and there are no obstacles which might obstruct the transmission. We assume the nodes are non-collinear and thus G_1 can have an arbitrary set of edges E_1 in the general case. As all the nodes use a fixed transmission range, the cost of the transmission is fixed. Let for simplicity $\omega(u, v) = 1$ for any pair of nodes. Then, $C(T, v) = cn(T, v)$ for any two nodes $u, v \in V$ and a convergecast tree T, i.e. the cost of a node is proportional to its congestion.

An example of this scenario is given in Figure 1. The input graph G_1 is shown in Figure 1a. The number in each node v represents $q(v)$. The resulting data gathering tree T is shown in Figure 1b. The number above each node represents its cost $C(T, v)$. Given that each node has the same initial battery power b, the lifetime of this network is equal to $\frac{b}{31}$.

In addition to arbitrary input graphs, we also consider an interesting case of k-layered graphs. A directed graph $H = (U, E_H)$ is a k-layered graph rooted at $r \in U$, if the nodes of the graph can be partitioned into k layers $U_1, U_2, \ldots, U_k$, such that $U_1 = \{r\}$ and E_H only contains inter-level edges from level i to level $i - 1$, for $2 \leq i \leq k$, i.e. $(u, v) \in E_H$ iff there exists i, $2 \leq i \leq k$ such that $u \in U_i$ and $v \in U_{i-1}$. An example of a 3-layered graph is shown in Figure 1a. Building the communication backbone in such graphs reduces



(a) Input Graph (b) The solution

Figure 1: The data gathering problem in the homogeneous scenario

system complexity and inefficient random access protocols impact, avoid inefficient mobile ad hoc routing, and alleviate the congestion bottleneck of the Internet gateway model [16].

We also pay special attention to 3-layered graphs. Such graphs are characterized by small number of hops, little interference, and are commonly used in wireless network design [22]. In addition, combing such networks with the OFDM mechanism provides the benefits of diversity and spatial reuse gain [36].

2.2.2 Heterogeneous network without obstacles

In heterogeneous networks, nodes can vary their transmission ranges, which directly affects the cost of the transmission. We use a common model for signal propagation [30] which states that $\omega(u, v) = d(u, v)^{\alpha}$, where α is the path-loss exponent parameter, which takes on a value between 2 and 4. We assume that $\alpha = 2$ for simplicity, however our results can be easily generalized for any value of α. As there are no obstacles the input graph G_2 for this scenario is the complete graph on the node set V.

3 Related work and our contribution

A number of recent papers has explored several issues and solutions related to energy efficient data gathering in wireless ad hoc networks under the homogeneous and heterogeneous models. In [2], Buragohain et al. show that the data gathering problem under the homogeneous model is $\mathcal{NP}$-hard when the initial battery charge at each node is different and have constant transmission power per message. They also proposed a $1 + c_r$ approximation algorithm for this problem (c_r is the cost of receiving a message), when node can aggregate incoming messages. Another model is by Wu et al. [38], who also study the data gathering problem with message aggregation and varying transmission power per message. They prove that this problem is $\mathcal{NP}$-hard and provide an $1 + \varepsilon$ approximation for it, for any $\varepsilon > 0$. A recent work by Liang et al. [24], explores the data gathering problem without aggregating messages at each node. They show a data gathering tree with lifetime that is $\Omega(\frac{\log n}{\log \log n})$ of the optimal, which means there is an instance of the optimal data gathering tree with lifetime better by a factor of at least $\frac{\log n}{\log \log n}$ with respect to their tree solution. In [17], Kalpakis et al. assume that the data at the sensors is highly correlated, and that the data gathering tree can be changed between transmission rounds. For this model they provide an integer program solution and

Approximation ratio	Model	Node layout	Remarks
Optimal (this paper)	homogeneous	$k = 3$	Each node has the same number of messages
2	homogeneous	$k = 3$	$\mathcal{NP}$-hard for $k \geq 3$ (this paper)
$\log n$	homogeneous	Arbitrary	Inapproximation ratio $\frac{\log n}{3}$ (this paper)
Optimal (this paper)	heterogeneous	Linear	$\mathcal{NP}$-hard for general graphs (this paper)
$\log q$ (this paper)	heterogeneous	Grid	
$O(\log^2 n \log q)$ (this paper)	heterogeneous	Grid	Nodes are uniformly distributed.

Figure 2: Summary of our results

a heuristic solution with no approximation ratios.

Liang et al. [27] show that the data gathering problem under the heterogeneous model it is $\mathcal{NP}$-hard. They provide several heuristic algorithms for this problem, and evaluated the performance of them through simulation. We note that their solution and hardness proof do not consider nodes' Euclidean position. For grid sensor networks, which we also investigate in this paper, Lin et al. [25] proposed a novel data gathering schema based on a chain-oriented grid architecture. They do not provide any analytical result for this model. Another interesting model for data gathering on grid sensor networks was studied by Bermond et al. [1]. They explore the data gathering problem with aggregation and node interference, where the optimization criteria is the number of hops until all messages are collected. For this model, they show a $1+\varepsilon$ approximation algorithm, for any $\varepsilon > 0$. For heterogeneous efficient data gathering in linear network, some work was done by Zhang et al. [40] and Liu et al. [26].

This paper is principally concerned with the theoretical and experimental study of the data gathering problem without aggregation under the homogeneous and heterogeneous models. Our main contributions are:

- Show that the problem is $\mathcal{NP}$-hard under both models. To the best of our knowledge, this is the first proof that considers the Euclidean positions of the nodes.
- Provide optimal solutions for the problem under special topologies such as 3-layered graphs and linear networks.
- Provide approximation algorithm for general graphs under both models.
- Verify by simulation that our results and models are correct.

Our results for each model are summarized in Figure 2.

4 Homogeneous networks with obstacles

In this section, we address the efficient data gathering problem on different topologies under the homogeneous model. This section is divided to k-layered graphs and general graphs. For 3-layered graphs, we start with

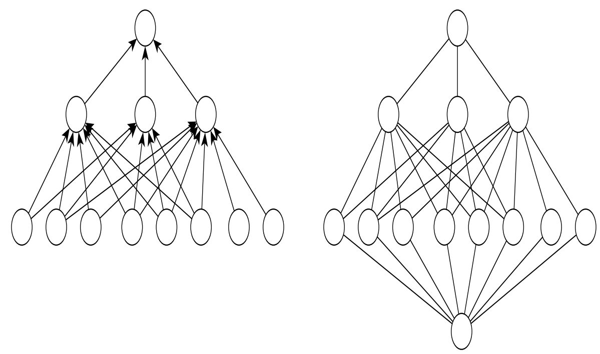


Figure 3: Optimal solution for the single message instance

a simple optimal solution when each node has only one message to transmit and provide a 2-approximation algorithm for the general case. We also show that this problem is $\mathcal{NP}$-hard when $k > 3$. For general graphs, we show that even if each node has only one message to transmit, it is polynomially hard to approximate the problem. Finally, we present an almost optimal approximation algorithm for general graphs with varying message size per node.

4.1 Layered Graphs

In the following, we show some results for the data gathering problem under the homogeneous model for layered graphs.

4.1.1 Single Message

Given an instance of the data gathering problem on a 3-layered graph, we create an auxiliary flow network by adding a new source node s, connecting it to all bottom nodes, setting the capacity of all edges to 1, and setting r to be the sink. We change the capacity of edges that connect intermediate nodes to the root from 1 to cp ($1 \leq cp \leq n$). Parameter cp represents the maximum congestion of the network. We search for a minimum cp, such that the maximum integer flow in the network is equal to n. This ensures that we can relay all n messages to the root, and that the maximum congestion is minimized. Thus this algorithm produce the optimal solution. Using binary search to find cp and Ford-Fulkerson algorithm [8] to find the maximum integer flow yields a running time of $O(nm \log n)$. The construction is illustrated in Figure 3. The input graph is shown on the left and the generated flow network on the right. Note that all arcs have capacity 1, except from arcs to the root with capacity cp. For this instance, setting cp=3 yields an optimal solution.

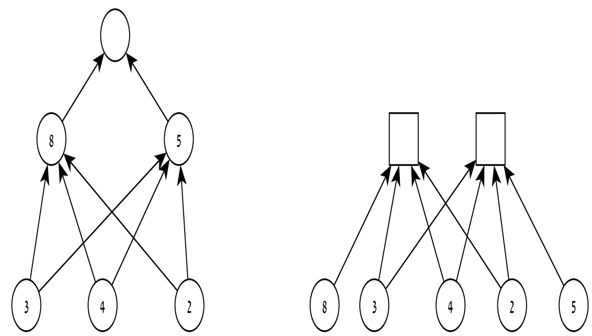


Figure 4: Reduction from the data gathering problem to SoUPM

4.1.2 Multiple Messages

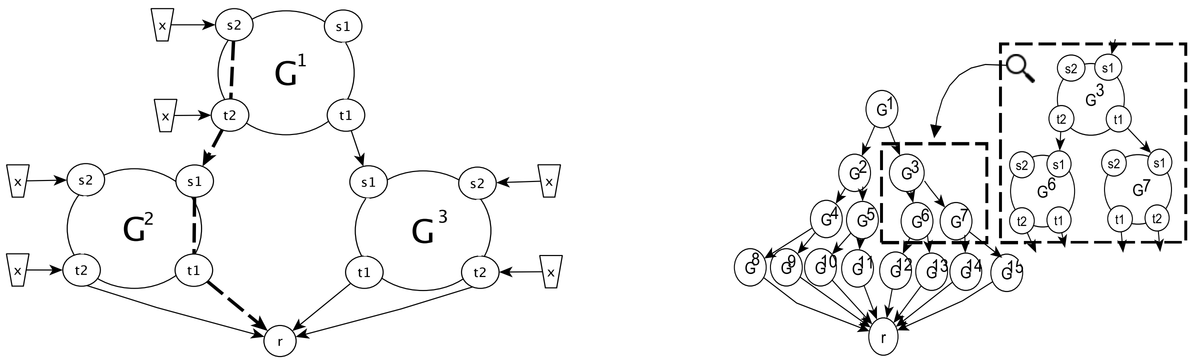
We show a reduction from the restricted assignment case of ***Scheduling on Unrelated Parallel Machines*** [21] (SoUPM) to the data gathering problem on 3-layered graphs with varying message size per node. In the restricted assignment case of SoUPM, n jobs must be assigned on m machines. The cost of assigning job j on specific machine is either p_j, the cost of the job, or ∞, which means job j cannot be assigned on that machine. The goal is to minimize the total cost of the most congested machine (also known as the makespan). Given an instance of SoUPM, we map the jobs to bottom nodes with weight p_j and the machines to intermediate nodes with weight 0. Since both problems have the same optimization criteria, the reduction shows that the data gathering problem is $\mathcal{NP}$-hard. A previous attempt to prove that the problem is $\mathcal{NP}$-hard can be found at [2]. However, the proposed reduction assumes varying battery size for each node. To transform an instance of the data gathering problem to SoUPM, we create a job per bottom node (with weight equal to the number of messages that node needs to transmit), and a machine per intermediate node. For intermediate nodes, we also add a dedicated job with weight equal to the node's messages. A sample reduction is depicted in Figure 4, the input graph is illustrated in the left side and the resulting instance of the SoUPM instance is illustrated on the right.

To approximate the data gathering problem, we transform the input graph to SoUPM, and use Lenstra et al. 2-approximation algorithm for SoUPM [21]. This leads us to the following lemma.

Lemma 4.1. *There is a* 2*-approximation algorithm for the data gathering problem on* 3*-layered graphs*

To show that the problem is $\mathcal{NP}$-hard for any $k > 3$, we reduce a 3-layered graph instance to a k-layered graph instance by:

1. Multiply the number of messages at each node in the 3-layered graph instance by $k - 2$. The cost of the solutions to the data gathering problems on a 3-layered graph is only multiplied by a constant (i.e., $C(T_{k-layer}, v) = (k - 2)C(T_{3-layer}, v)$).



(a) Auxiliary graph G for the case when $n = 4$ (b) Underlying Tree structure T

Figure 5: The auxiliary graph $\tilde{G}$

2. Transform all bottom nodes (nodes in layer-3) to a directed path with length $k - 2$, and split the messages of the bottom nodes equally between them.

Clearly, we can use the k-layered graph algorithm to solve the 3-layered graph instance. Thus, the following lemma holds:

Lemma 4.2. *The data gathering problem is $\mathcal{NP}$-hard on k-layered graphs for any $k > 3$.*

4.2 General graphs

We show that the data gathering problem is $\mathcal{NP}$-hard under the homogeneous model for general graphs and that it cannot be polynomially approximated with a ratio better than $\frac{\log n}{3}$. We also show how to achieve a $O(\log n)$ approximation algorithm for it.

4.2.1 Inapproximation

We prove that the data gathering problem cannot be polynomially approximated by a ratio better than $\frac{\log n}{3}$, unless $\mathcal{P} = \mathcal{NP}$, even if each node has only one message to transmit.

We use a) Gadget by Guruswami et al. [13] for the edge-disjoint path problem and b) the underlying tree structure connecting gadgets by Chen et al. [4]. We also use the $\mathcal{NP}$-hard 2-*vertex disjoint path problem* [9], where we are given a directed graph G, and four special nodes s_1, s_2, t_1, t_2, and we need to decide if G contains 2-vertex-disjoint directed paths from s_1 to t_1 and from s_2 to t_2. Given an instance $\langle G, s_1, s_2, t_1, t_2 \rangle$ of the 2-vertex disjoint directed paths problem, we create a complete binary tree T with $\log n$ levels (n is the number of nodes in G). Notice that each level in T contains twice as much nodes as the previous one. We use T to create an auxiliary graph $\tilde{G}$ by transforming every node in T to a copy of G (the original graph). We denote G^i and G^i_v as the i^{th} copies of G, and a copy of the node v in G^i, respectively. We connect the nodes from level l ($l \geq 1$) of form $G^i_{t_2}$ ($2^{l-1} \leq i < 2^l$) to nodes in level $l+1$ of form $G^{2i}_{s_1}$ and $G^{2i+1}_{s_1}$, respectively. We attach a directed path with $n \log n$ nodes to each node with the form $G^i_{t_2}$, $G^i_{s_2}$. Finally, we create a sink node r and connect each node that has the form $G^i_{t_1}$, $G^i_{t_2}$ and located in level $\log n$ to it ($2^{\log n - 1} \leq i < 2^{\log n}$-1).

An example of the construction is shown in Figures 5(a) and 5(b). In Figure 5(a) we have $n = 4$ with 3 copies (x denotes a directed path having $n \log n$ nodes). The connections between the levels are $G^1_{t_1} \rightarrow G^3_{s_1}$ and $G^1_{t_2} \rightarrow G^2_{s_1}$. $G^2_{t_1}, G^2_{t_2}, G^3_{t_1}, G^3_{t_2}$ are connected to the sink.

Denote the cost of the optimal solution for the data gathering problem on a graph $\tilde{G}$ as OPT. We state the following:

Lemma 4.3. *If G contains 2 vertex disjoint directed paths, then $OPT < 3n \log n$.*

Proof. First note that any path to the sink that starts from nodes with form $G^i_{s_1}$ will have at most $n \log n$ nodes (adding at most n nodes to every $G^i_{t_1}$in each level). Also note that we can create a directed path from every node $G^i_{s_2}$ to the sink with at most $3n \log n$ nodes (the path from $G^i_{s_2}$ to $G^i_{t_2}$ will have at most $2n \log n + n$ nodes, and at another $(\log n - 1)n$ node in the path to r). Thus, if G contains 2 vertex disjoint directed paths the maximum cost of a node $\max C(T, v)$ is less than $3n \log n$. Hence, $OPT < 3n \log n$.

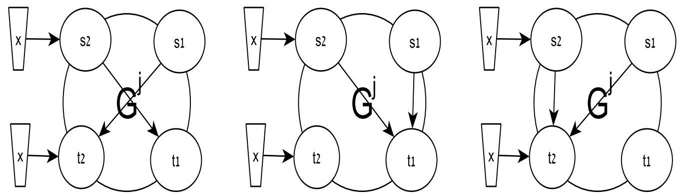
A sample path $G^1_{s_2} \rightsquigarrow G^1_{t_2} \rightarrow G^3_{s_1} \rightsquigarrow G^3_{t_1} \rightarrow r$ is shown by bold dashed lines in Figure 5a. □

Lemma 4.4. *If G does not contains 2 vertex disjoint directed paths, then $OPT > n \log^2 n$.*

Proof. First, we prove that if the maximum cost of a node in level i is $c \cdot n \log n$ (for any constant c) and G does not contain 2 -vertex disjoint directed paths, then there is a node in level $i + 1$ with cost at least $(c + 1) \cdot n \log n$. We prove this using induction on the number of levels. The base case is obvious (in level 1 to cost of $G^1_{s_2}$ is $n \log n$). For the induction step, suppose that there exists a node G^l_v in level i with cost $c \cdot n \log n$. This means that one of the nodes $G^j_{s_1}$ in level $i + 1$ will cost at least $c \cdot n \log n$. Then, since we do not have disjoint paths, we face one of the cases depicted in Figure 6. In each case, either $G^j_{t_1}$ or $G^j_{t_2}$ will cost at least $(c + 1) \cdot n \log n$. This implies that the induction hypothesis holds, and since we have $\log n$ levels, the maximum cost of a node is at least $n \log^2 n$. □

Theorem 4.5. *The data gathering problem cannot be polynomially approximated within a factor better than $\frac{\log n}{3}$, unless $\mathcal{P} = \mathcal{NP}$.*

Proof. Lemma 4.3 implies that if G contains 2-vertex disjoint directed paths then $OPT < 3n \log n$. Lemma 4.4 implies that if G does not contain 2 vertex disjoint directed paths, then $OPT > n \log^2 n$. We note that the gap between the instances is $\frac{n \log^2 n}{3n \log n} = \frac{\log n}{3}$. Now, suppose there exists a polynomial algorithm $\mathcal{A}$ that approximate data gathering problem with approximation ratio better than $\frac{\log n}{3}$. If G contains 2-vertex disjoint paths, $\mathcal{A}$ will produce a solution with cost at most $\frac{\log n}{3} \cdot 3n \log n = n \log^2 n$. Otherwise, the solution



(a) case (i) (b) case (ii) (c) case (iii)

Figure 6: No 2-disjoint paths

$\mathcal{A}$ produce has cost at least $n \log^2 n$ $(OPT > n \log^2 n)$. Consequently, using $\mathcal{A}$, we can decide if G contains 2-vertex disjoint sets or not. Thus, unless $\mathcal{P} = \mathcal{NP}$ the data gathering problem cannot be polynomially approximated by a ratio better than $\frac{\log n}{3}$. □

4.2.2 A $O(\log n)$ Approximation Algorithm

First, we define the notion of a *confluent flow*. Given a general network flow problem, in a *confluent flow*, all the out flow per node must leave along a single edge, i.e., the resulting flow network is a tree. To solve the data gathering problem, we use Chen et al. CONFLT algorithm [4], which achieves a $O(\log n)$ approximation algorithm for the single commodity confluent flow problem.

The input for the CONFLT algorithm is a graph G, set of sinks s, set of demands d, and a splittable flow f. The output of the algorithm is a confluent flow $\widetilde{f}$ with tree topology. We refer to the maximum outgoing flow per node as congestion (recall that in the data gathering problem, congestion is the number of messages each node delivers, which maps to the total amount of out flow per node in this case). The CONFLT algorithm guarantees that: a) the flow conservation constraints hold in $\widetilde{f}$, b) the outgoing flow from each node will leave along a single edge, and c) if the maximum congestion of f is 1, then the maximum congestion of $\widetilde{f}$ is $1 + \log n$.

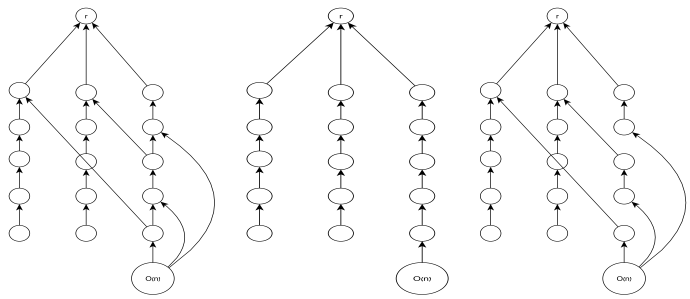
Our algorithm steps are as follows: given a data gathering instance G and a root r, we:

1. Create a flow network on the underlying graph G, and define the demand per node as the amount of messages this node needs to transmit.

2. search for a splittable flow with minimal congestion that serves all demands (i.e., all messages are delivered).

3. Scale the resulting splittable flow to 1 and run CONFLT on it.

The running time of combing CONFLT and the maximum flow algorithm by King et al. [18] is
$O(mn(\log n \log_{\frac{m}{n \log n}} n + \frac{m}{n} + \log \frac{n^2}{m}))$.



(a) Original graph (b) ECRT solution (c) CONFLT solution

Figure 7: Graph instance where ECRT approximation ratio is $\Omega(n)$ while CONFLT is optimal

4.2.3 Comparison to other data gathering algorithm

To emphasize the efficiency of CONFLT this subsection includes a theoretical comparison to other existing techniques of lifetime maximization in data gathering without aggregation. To solve the data gathering problem, other schemes usually use heuristics, which are based on a greedy approach of constructing the data gathering tree. For example, ECRT [2] greedily grows the tree by appending nodes which minimize the local lifetime, and MITT [24] constructs a minimum spanning tree and then greedily replaces the edges to increase the network lifetime.

It turns out that there are some graph instances where the approximation ratio of such algorithms is $\Omega(n)$, while the approximation ratio of CONFLT is $O(\log n)$ or better. For example, in Figure 7(a), we have a graph where the root is connected to 3 children, each connected to a chain of nodes, with one child that has a chain that ends with a very large cluster of nodes (the size of the cluster is in the order of n). The cluster is only connected to 3 nodes in the rightmost chain, and each of those nodes is connected to a a different child of the root.

The result of running ECRT is depicted in Figure 7(b). Since the algorithm is greedy, the nodes in the large cluster are relayed only by the rightmost child of the root. Therefore, the maximum node's cost $C(T, v)$ is approximately n, the size of the large cluster. CONFLT result is depicted in Figure 7(c). In this algorithm, the load is distributed between all three children of the root (i.e., the maximum cost is approximately $\frac{1}{3}n$). Extending this graph structure to include m children will yield the desired approximation ratios, i.e., ECRT cost will be $\Omega(n)$ while CONFLT cost will be $\frac{n}{m}$.

5 Heterogeneous networks without obstacles

In this section, we investigate the algorithmic complexity of the data gathering problem under the heterogeneous model. In this model, the nodes are deployed in the Euclidean plane, and the cost of sending a

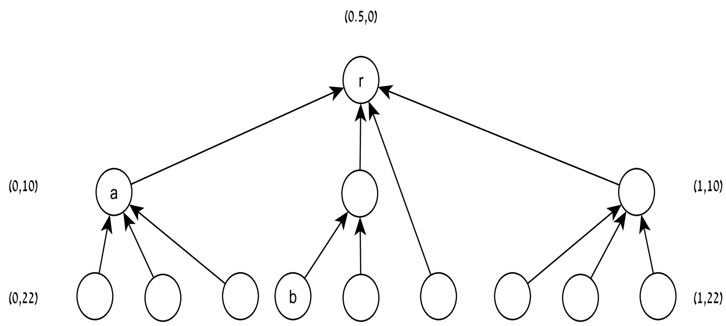


Figure 8: Reduction from 3-partition

message is $d(u,v)^\alpha$ (where $d(u,v)$ is the Euclidean distance between nodes u and v, and α represents the path-loss exponent). To simplify the analysis, the results in this sections are shown for $\alpha = 2$, but can be extended for any $\alpha > 2$. . First, we prove that the data gathering problem is $\mathcal{NP}$-hard. Next, we show how to solve the data gathering problem for three special instances, linear networks, where the nodes are placed on a bi-directional line, static grids topologies, and random grids, where the nodes are uniformly distributed on the surface of an $\sqrt{n} \times \sqrt{n}$ square. We explore the results for the static and random grids using simulation in Section 6.

5.1 $\mathcal{NP}$-hardness of the data gathering problem

We prove that the data gathering problem is $\mathcal{NP}$-hard in the strong sense using a reduction from the extended version of the 3-partition problem. In this problem, we are given a multi-set S of $3m$ elements $S = \{a_1, a_2, .., a_{3m}\}, a_i \in \mathbb{Z}^+$, where each element has a weight strictly between $\frac{B}{4}$ and $\frac{B}{2}$, and all the elements have a total weight of mB. The goal is to decide whether there is a partition of S into m sets each of weight B, such that the union of those sets covers S (each set contains exactly 3 elements) [11]. For the reduction, we use the decision version of the data gathering problem, where we are given an instance graph G_E (a complete undirected Euclidean graph), and an integer P. The goal is to decide whether there is a data gathering tree T, where the maximum node cost $C(T,v)$ is less than or equal to P.

Let $I = \langle S \rangle$ be an instance of the extended 3-partition problem. First, we create a mapping between S to G_E, the input graph for the data gathering problem. For each element a_i, in the partition instance, we create a node v_i with a_i messages to transmit, and position those nodes between coordinates $(0,22)$ to $(1,22)$ (with equal distance between each node).

Next, we create m intermediate nodes $s_1, s_2, ..s_m$, each with one message to transmit, and position them between coordinates $(0,10)$ to $(1,10)$. Those nodes correspond to the m bins of the partition instance. Finally, the root is placed at coordinates $(0.5,0)$. The reduction is depicted in Figure 8. Note that the proof still holds even if the Euclidean positions of the nodes scale (for example, by multiplying all x and y coordinates by a factor of 10).

Using the construction, we claim the following:

Lemma 1. *If a solution to the data gathering problem with a cost less than or equal to* $(B+1)(10^2+0.5^2)$ *exists, than each intermediate node will carry exactly* 3 *nodes from the lower layer with total message cost* B.

Proof. Suppose such solution exists, we make the following observation:

1. Sending a message from a node v_i (node that corresponds to a partition element) to the root will cost at least:
$$(22)^2\frac{B}{4} = 121B$$
We infer that in the tree solution such nodes will not deliver their messages directly to the root (i.e., they will use intermediate nodes).

2. Relaying the messages from one of the bottom nodes to one of the intermediate nodes costs at most:
$$(12^2+1^2)\cdot\frac{B}{2} = 72.5B$$
We infer that in the tree solution the node with the maximum cost $\max C(T,v)$ is an intermediate node.

3. The total number of messages at the bottom nodes is mB, and they are divided between m intermediate nodes (which relay the messages to the root). From the pigeonhole principle, each intermediate node will relay messages from 3 nodes that have a total of B messages.

By combing the above arguments, we get that in a solution with cost less than or equal to $(B+1)(10^2+0.5^2)$, each intermediate node relay $B+1$ messages (including its own message), and only intermediate nodes communicate with the root. Note that since the maximum distance, $\sqrt{10^2+0.5^2}$, is between the leftmost and rightmost nodes to the root, the total cost of the solution is exactly $(B+1)(10^2+0.5^2)$.

□

If a solution with cost $(B+1)(10^2+0.5^2)$ exists, we can find a solution to the extended 3-partition problem by joining each triplet a_i, a_j, a_k with total messages weight B that uses intermediate node s_l as a carrier.

Hence, unless $\mathcal{P}=\mathcal{NP}$ we get the following theorem:

Theorem 5.1. *The data gathering problem is $\mathcal{NP}$-hard under the Heterogeneous model.*

Additional attempt to prove that the problem is $\mathcal{NP}$-hard can be found at [27], note that the proposed reduction does not consider the Euclidean positions of the nodes.

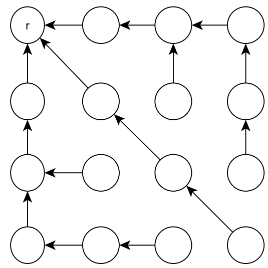


Figure 9: Sample solution for the 4×4 grid

5.2 Line topology

We show the optimal solution for the data gathering problem on a line topology with n equally placed nodes is a directed line (chain) from the fringe nodes to the root. Since the problem is symmetric, w.l.o.g we locate the root in the left most corner of the graph. We order the nodes from node v_n at location n to the root at location 0. We state the following theorem:

Theorem 5.2. *If the data gathering tree is not a directed line, some messages will not be delivered.*

Proof. Let v_j be the rightmost neighbor of v_1 that delivers a direct message to r, it must be that $j \leq \sqrt{n}$ (otherwise the cost of this node will be at least $j^2 > n$). This implies that the maximum number of messages that node v_j can pass is $\left\lfloor \frac{n}{j^2} \right\rfloor$. Observe that if node v_j passes a message directly to the root, then node v_1 can only deliver $\sum_{i=2}^{n} \left\lfloor \frac{n}{i^2} \right\rfloor + j - 1$ messages to r ($\sum_{i=2}^{n} \left\lfloor \frac{n}{i^2} \right\rfloor$ corresponds to the total number of messages that bypass node v_j and $j-1$ corresponds to the left neighbors of v_j). The total number of messages that can be delivered to r this way are:

$$\sum_{i=2}^{\sqrt{n}} \left\lfloor \frac{n}{i^2} \right\rfloor + j - 1 + \left\lfloor \frac{n}{j^2} \right\rfloor \leq \sum_{i=2}^{\infty} \frac{n}{i^2} + j + \frac{n}{j^2} \leq n(\frac{\pi^2}{6} - 1) + \frac{n}{3^2} + \sqrt{n}.$$

The inequalities derive from the fact that $\sum_{i=2}^{\infty} \frac{n}{i^2} = n(\frac{\pi^2}{6} - 1)$ and that $3 \leq j \leq \sqrt{n}$. For $n \geq 20$, we get that this is less or equal to $n(0.65 + 0.12 + 0.225)$, which is less than n, the amount of messages that must be delivered to the root. Therefore, any data gathering solution with cost less than n is not a valid solution since not all messages arrive to the root. □

In contrast, the cost of a directed line is exactly n and all n messages are delivered. Hence, it is the optimal solution.

5.3 Grid topologies

The input for the problem is a complete graph with n nodes located on the $\sqrt{n} \times \sqrt{n}$ grid, where the root r is located in the lower left corner of the grid at coordinates $(0, 0)$. We denote q as the total messages from all nodes. We show that on this topology the cost of any data gathering tree rooted at r is at least $\frac{q}{5\log n}$. Then we present a deterministic construction of tree T having cost $\frac{q}{2}$, implying a $\log n$ approximation solution for this problem.

Lemma 5.3. *The cost of any data gathering rooted at r is at least $\frac{q}{3\log n}$.*

Proof. Denote by $d_{i,j}$ the Euclidean distance from the node located at coordinate (i, j) of the grid to r, and by $d_{i,j}^2$ the cost of sending a direct message to r. Assume we have an optimal tree solution for the problem with cost p^*, where $p^* \le q$. Thus, every node located at (i, j) can relay only $\frac{p^*}{d_{i,j}^2}$ messages to the root. The total number of messages that can be delivered to r, keeping the cost below p^*, is:

$$\sum_{i,j:\sqrt{i^2+j^2}\le\sqrt{p^*}} \frac{p^*}{d_{i,j}^2} \le \sum_{i,j:\sqrt{i^2+j^2}\le\sqrt{p^*}} \frac{p^*}{d_{i,1}^2} \le p^* \sum_{i=1}^{\sqrt{p^*}} (2i+1)\frac{1}{i^2}$$

The first inequality derives from the fact that $\frac{1}{d_{i,j}^2} \le \frac{1}{d_{i,1}^2}$ for any $j \in \mathbb{N}$. The second inequality derives from the fact that $d_{i,1} = i$ and that for each i, there are exactly $2i+1$ nodes with either $x = i$ or $y = i$ coordinates (e.g., for $i = 1$, we have the nodes located at $(0, 1), (1, 0)$ and $(1, 1)$). Replacing the summation using the generalized harmonic number ($H_{n,m} = \sum_{k=1}^{n} \frac{1}{k^m}$) we get:

$$p^*(2H_{\sqrt{p^*},1} + H_{\sqrt{p^*},2}) \le p^* 3 \log p^* \le p^* 3 \log q$$

Since $H_{n,1} \sim \log n + \frac{1}{2n} + \gamma$ ($\gamma \sim 0.577$ is the Euler-Mascheroni constant), we can replace $H_{\sqrt{p^*},1}$ by $2 \log \sqrt{p^*}$ and $H_{\sqrt{p^*},2}$ by $\log \sqrt{p^*}$. In addition, according to the assumption $p^* \le q$.

Since q messages are delivered to r, $q \le p^* 3 \log q$. Hence, p^* is at least $\frac{q}{3\log q}$. □

Theorem 5.4. *There is a solution with cost at most $\sqrt{2}q$.*

Proof. Passing a directed line through the diagonal nodes directly to the root while moving all the other nodes through the side nodes yields a solution with maximum cost of $\sqrt{2}q$ (see Figure 9 for an example on the 4×4 grid). Thus, this algorithm is a $\log q$ approximation for the data gathering problem on the $\sqrt{n}\times\sqrt{n}$ grid. □

Note that the minimum spanning tree (MST) on any $\sqrt{n} \times \sqrt{n}$ grid has maximum unit edge cost. Since the maximum cost of a node $\max C(T, v)$ in the MST will be equal to $1^2(q-1)$, MST also presents a $O(\log n)$ approximation algorithm.

5.4 Uniformly distributed nodes

We show how to achieve a $O(\log^2 n \log q)$ approximation algorithm for the data gathering problem when nodes are uniformly distributed in the $\sqrt{n} \times \sqrt{n}$ square U.

We state the following:

Lemma 5.5. *For n nodes uniformly distributed in a $\sqrt{n} \times \sqrt{n}$ square U , if we divide U into a $\sqrt{\frac{n}{\log n}} \times \sqrt{\frac{n}{\log n}}$ grid with equal size cells, then w.h.p. (with high probability) each cell will contain at least 1 node[33].*

Lemma 5.6. *For n nodes uniformly distributed in a $\sqrt{n} \times \sqrt{n}$ square U, if we divide U into a $\sqrt{\frac{n}{\log n}} \times \sqrt{\frac{n}{\log n}}$ grid with equal size cells, then w.h.p. each cell will contain at most $e^3 \log n$ nodes.*

Proof. The proof is straightforward from the result of *Balls and Bins* [19]. □

Theorem 5.7. *For n nodes uniformly distributed in a unit square U, w.h.p. there is an $O(\log^2 n \log q)$ approximation algorithm for the data gathering problem.*

Proof. We start by emulating a $\sqrt{\frac{n}{\log n}} \times \sqrt{\frac{n}{\log n}}$ grid on the plane. From the previous results (Lemma 5.5 and Lemma 5.6) w.h.p the number of nodes in each cell is between one to $\alpha \log n$ (where α is a constant). We define $|V_{i,j}|$ as the number of nodes in the cell with coordinate (i, j). Counting the total number of messages we can deliver to r, when $\max C(T, v) \leq p^*$, using the same reasoning as in Lemma 5.3 we get:

$$\sum_{i,j:\sqrt{i^2+j^2}<\sqrt{p^*}} \frac{p^*}{d_{i,j}^2} \cdot |V_{i,j}| < \sum_{i=1}^{\sqrt{p^*}} (2i+1) \frac{p^*}{d_{1,i}^2} \cdot \alpha \log n \leq p^* \alpha \log n \sum_{i=1}^{\sqrt{p^*}} (2\frac{1}{i} + \frac{1}{i^2}) \leq p^* 3\alpha \log n \log p^*$$

Since $p^* \leq q \log n$ and $n \leq q$, the expression is less than or equal to:

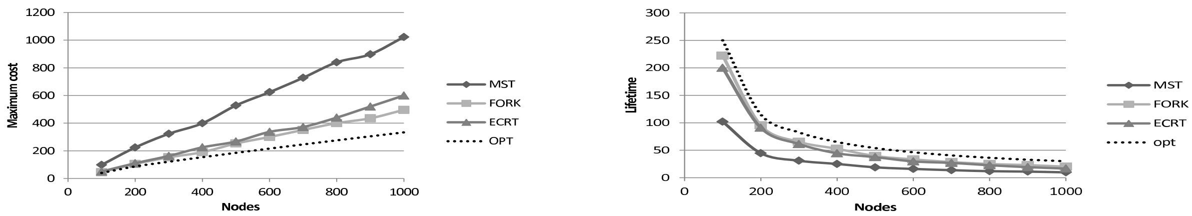
$$p^* 3\alpha \log n \log (q \log n) \leq p^* 3\alpha \log n \log q^2 \leq 6\alpha \log n \log q$$

Finally, since q message are delivered to the root, the total invested energy p^* is at least $\frac{q}{6\alpha \log n \log q}$.

To solve the data gathering problem, we create a minimum spanning tree MST. If all the nodes in the network transmit with power $\log n$, then w.h.p the network will became connected [14]. This implies that the transmission cost of the MST is at most $\log n$. Thus, the maximum cost at each node $C(T, v)$ is at most $q \log n$, which yields a $O(\log^2 n \log q)$ approximation for the data gathering problem. For a distributed implementation of this algorithm, see [10]. □

6 Simulation study

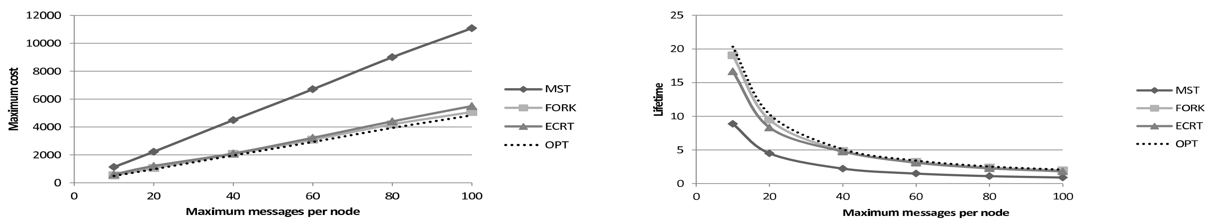
In this section, we provide performance evaluation results for the data gathering problem under the homogeneous and heterogeneous models (which were studied in Sections 4 and 5, respectively). In all simulations, each node has an initial battery capacity of 10000. We compare the results of running algorithms presented in this paper, to the theoretical optimum, and the ECRT heuristic from [2]. In the ECRT heuristic, which provides good simulation results for the data gathering problem under the homogeneous and the heterogeneous models, the algorithm greedily increases the size of the data gathering tree by adding edges that have minimum impact on the lifetime of the network. We refer to the algorithm for static grid from Section



(a) Maximum cost of a node - $C(T, v)$

(b) Lifetime - $L(T)$

Figure 10: Static grid - network size varies



(a) Maximum cost of a node - $C(T, v)$

(b) Lifetime - $L(T)$

Figure 11: Static grid - number of messages per node varies

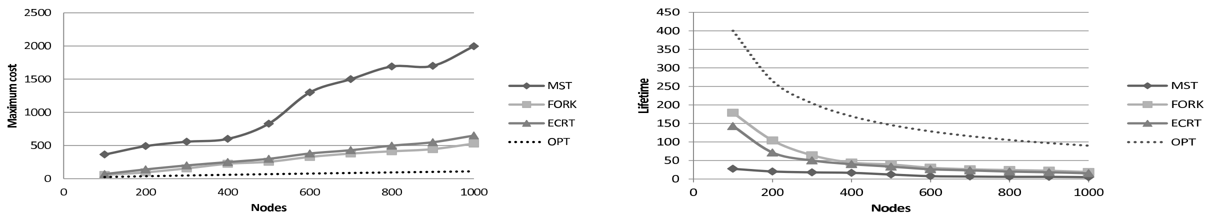
5.3 and for random grid from Section 5.4 as FORK, since the resulting tree has a fork-like structure. All algorithms were implemented in .NET environment and run on a common desktop PC (Dell Optiplex 990). In our simulations, the network size varies from 100 nodes to 1000 nodes while the number of messages is in the interval 1 to 100. Each point the plots is an average of 10 trials. We have considered different scenarios for homogenous and heterogeneous networks with fixed and various number of messages.

6.1 Heterogeneous model - static grid

For each simulation, we position n nodes inside a $\sqrt{n} \times \sqrt{n}$ square; every node has one message to transmit. The communication infrastructure is a complete graph and the cost of sending a message is equal to $d(u, v)^2$. We compared the performance of the minimum spanning tree (MST), our algorithm from Section 5.3 (FORK) and the ECRT heuristic.

The cost of the maximum node, $\max C(T, v)$, for MST, FORK and ECRT is plotted in Figure 10(a). The results reinforce the theoretical findings. For MST, $C(T, v)$ is approximately n (since the root has only one child), and for FORK and ECRT, $C(T, v)$ is approximately $\frac{1}{2}n$. From the graphs, we can see that FORK provides better results than ECRT. The lifetime graph is plotted in Figure 10(b). Given that CPU and power optimization are a major criteria when designing ad-hoc networks, FORK performance is better than ECRT on static grids.

The total cost and lifetime graphs for 200 nodes network with varying message size are plotted in Figures 11(a) and 11(b), respectively. From the graphs, we learn that the total cost increases linearly with the



(a) Maximum cost of a node - $C(T, v)$

(b) Lifetime - $L(T)$

Figure 12: Random grid - network size varies

average number of messages per node $\bar{q}$, note that in FORK, $C(T, v)$ is approximately $\frac{\bar{q}}{2}$. Again, FORK slightly outperforms both MST and ECRT.

6.2 Heterogeneous graphs - random grid

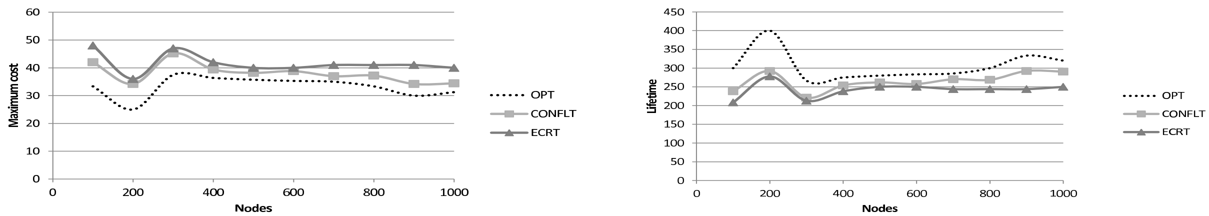
For every simulation, the nodes were randomly, independently and uniformly distributed in a $\sqrt{n} \times \sqrt{n}$ square, each with one message to transmit. The evaluation of the maximum node energy cost $C(T, v)$ of MST, FORK and ECRT is plotted in Figure 12(a). The cost is compared to the theoretical optimum, $\frac{n}{\log^2 n}$, which we proved in Section 5.4. As expected, the cost of MST has a polylogarithmic dependency in the number of nodes (i.e., $C(T, v) = OPT \cdot O(\log^2 n)$). Note that both FORK and ECRT maximum cost is about $\frac{n}{2}$ (which may imply their approximation ratio is much better than $O(\log^2 n)$.

The lifetime of the network is depicted in Figure 12(b). Overall, both FORK and ECRT yield good results both in lifetime and in maximum cost, with a slight advantage to FORK. However, ECRT is not practical on random networks since it requires the centralized processing (all nodes must have complete network information) and its processing time to construct the tree is $O(n^4)$ for complete graphs. This makes ECRT extremely costly and unpractical in both CPU usage and power consumption. In contrast, FORK scheme is completely distributed, and its processing time to construct the data gathering tree is $O(\log n)$. Given that most ad-hoc networks suffer from CPU and network bandwidth constraints, the fast and distributed implementation of FORK is preferable.

6.3 Homogeneous graphs

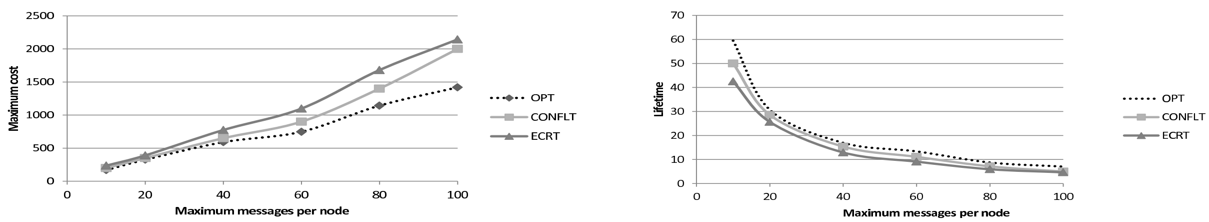
For all simulations, we randomly scattered the nodes over a 1000×1000 square meters area, and formed a unit disk graph (UDG) by assigning each node a transmission radius of 100 meters.

We compared the results of algorithm CONFLT from Section 4.2.2, ECRT heuristic, and the theoretical lower bound on OPT, which is defined as n divided by the number of root's children. The maximum node cost $C(T, v)$ is plotted in Figure 13(a). We observe an interesting phenomenon that the cost for both algorithms stabilizes as n increases. This can be explained by the fact that although there are more nodes, the ratio between the number of root's children and total number of nodes in the network does not change much, and



(a) Maximum cost of a node - $C(T, v)$ (b) Lifetime - $L(T)$

Figure 13: Homogeneous graphs - number of nodes varies



(a) Maximum cost of a node - $C(T, v)$ (b) Lifetime - $L(T)$

Figure 14: Homogeneous graphs - number of messages per node varies

therefore the relative load of every node remains almost the same. From the results, we can learn that the approximation ratio is constant for both algorithms (i.e., $O(1)$ from OPT). The lifetime of the network is plotted in Figure 13(b).

For the case, when the number of messages per node varies, the maximum node's cost and lifetime graphs, are depicted in Figures 14(a) and 14(b), respectively. Similarly to the results for static grid (Figure 11), the total cost increases linearly with the average number of messages per node $\bar{q}$. To conclude, for all simulations, the lifetime and maximum load achieved by CONFLT is better.

7 Conclusions and future work

In this paper, we have studied the problem of constructing energy efficient non-aggregated data gathering tree under the homogeneous and heterogeneous network models. We have shown the problem is $\mathcal{NP}$-hard under both models, and provided several approximation algorithms for it. We also provided simulations that support the theoretical results for the heterogeneous model. The implementation of some of the algorithms depends on a centralized base station that perform the calculation and power assignment. In future work, we plan to investigate a distributed scalable solution for the problem under the lifetime optimization metric. Another future research direction is data gathering in complex evolving environment, when nodes' location, number of messages, or battery power changes over time.

Acknowledgments The authors would like to thank Refael Hassin and the anonymous reviewers for their valuable suggestions and helpful comments.

References

[1] J. C. Bermond, N. Nisse, P. Reyes, and H. Rivano. Fast data gathering in radio grid networks. In *AlgoTel'09*, pages 53–56, 2009.

[2] C. Buragohain, D. Agrawal, and S. Suri. Power aware routing for sensor databases. In *INFOCOM 2005*, pages 1747–1757, 2005.

[3] A. Chehri, P. Fortier, and P.-M. Tardif. Security monitoring using wireless sensor networks. *Communication Networks and Services Research, Annual Conference on*, 0:13–17, 2007.

[4] J. Chen, R. D. Kleinberg, L. Lovász, R. Rajaraman, R. Sundaram, and A. Vetta. (almost) tight bounds and existence theorems for single-commodity confluent flows. *J. ACM*, 54, July 2007.

[5] T.-S. Chen, H.-W. Tsai, and C.-P. Chu. Gathering-load-balanced tree protocol for wireless sensor networks. pages 8–13, 2006.

[6] K. Chintalapudi, T. Fu, J. Paek, N. Kothari, S. Rangwala, J. Caffrey, R. Govindan, E. Johnson, and S. Masri. Monitoring civil structures with a wireless sensor network. *IEEE Internet Computing*, 10(2):26–34, 2006.

[7] D. Ciullo, G. D. Celik, and E. Modiano. Minimizing Transmission Energy in Sensor Networks via Trajectory Control. In *WiOpt'10: Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, pages 259–268, 2010.

[8] L. R. Ford and D. R. Fulkerson. Maximal flow through a network. *Canadian Journal of Mathematics*, 8:399–404, 1956.

[9] Fortune, S. Hopcroft, J. E., Wyllie, and J. C. The directed subgraph homeomorphism problem. *Theoret. Comput. Sci.*, 10:111–121, 1980.

[10] R. G. Gallager, P. A. Humblet, and P. M. Spira. A distributed algorithm for minimum-weight spanning trees. *ACM Trans. Program. Lang. Syst.*, 5:66–77, January 1983.

[11] M. R. Garey and D. S. Johnson. *Computers and Intractability; A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., New York, NY, USA, 1990.

[12] J. Gehrke and S. Madden. Query processing in sensor networks. *IEEE Pervasive Computing*, 3(1):46–55, 2004.

[13] V. Guruswami, S. Khanna, R. Rajaraman, B. Shepherd, and M. Yannakakis. Near-optimal hardness results and approximation algorithms for edge-disjoint paths and related problems. *J. Comput. Syst. Sci.*, 67(3):473–496, 2003.

[14] R. Hekmat and P. Van Mieghem. Connectivity in wireless ad-hoc networks with a log-normal radio model. *Mob. Netw. Appl.*, 11:351–360, June 2006.

[15] J. Hoebeke, I. Moerman, B. Dhoedt, and P. Demeester. An Overview of Mobile Ad Hoc Networks: Applications and Challenges. *Journal of the Communications Network*, 3:60–66, July 2004.

[16] Hung-yu Wei and R. D. Gitlin. Two-hop-relay architecture for next generation wwan/wlan integration. *IEEE Wireless Communications (Special Issue on 4G Mobile Communications - Towards Open Wireless Architecture),*, 11:24–30, April 2004.

[17] K. Kalpakis, K. Dasgupta, and P. Namjoshi. Efficient algorithms for maximum lifetime data gathering and aggregation in wireless sensor networks. *Computer Networks*, 42:697–716, 2003.

[18] V. King, S. Rao, and R. Tarjan. A faster deterministic maximum flow algorithm. pages 157–164, 1992.

[19] V. F. Kolchin, B. A. Sevastianov, and V. P. Chistiakov. *Random allocations*. V. H. Winston ; distributed solely by Halsted Press, Washington : New York :, 1978.

[20] J. Kulik, W. Heinzelman, and H. Balakrishnan. Negotiation-based protocols for disseminating information in wireless sensor networks. *Wirel. Netw.*, 8:169–185, March 2002.

[21] J. K. Lenstra, D. B. Shmoys, and É. Tardos. Approximation algorithms for scheduling unrelated parallel machines. *Math. Program.*, 46(3):259–271, 1990.

[22] J. Lessmann and A. Krishnamurthy. Parameterized hierarchical layer topology construction for wireless networks. In *ICSNC '07: Proceedings of the Second International Conference on Systems and Networks Communications*, page 15. IEEE Computer Society, 2007.

[23] L. Levin, M. Segal, and H. Shpungin. Optimizing Performance of Ad-hoc Networks Under Energy and Scheduling Constraints. In *WiOpt'10: Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, pages 110–119, 2010.

[24] J. Liang, J. Wang, J. Cao, J. Chen, and M. Lu. An efficient algorithm for constructing maximum lifetime tree for data gathering without aggregation in wireless sensor networks. In *Proceedings of the 29th conference on Information communications*, INFOCOM'10, pages 506–510. IEEE Press, 2010.

[25] C.-S. Lin, C.-N. Huang, and R.-J. Fang. A power-efficient data gathering scheme on grid sensor networks. In *Proceedings of the 8th WSEAS International Conference on Multimedia systems and signal processing*, pages 142–147. World Scientific and Engineering Academy and Society (WSEAS), 2008.

[26] A.-F. Liu, X.-Y. Wu, Z.-G. Chen, and W.-H. Gui. An energy-balanced data gathering algorithm for linear wireless sensor networks. *International Journal of Wireless Information Networks*, 17:42–53, 2010.

[27] Y. Liu. Online data gathering for maximizing network lifetime in sensor networks. *IEEE Transactions on Mobile Computing*, 6(1):2–11, 2007.

[28] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson. Wireless sensor networks for habitat monitoring. In *WSNA '02: Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, pages 88–97, 2002.

[29] V. Mhatre and C. Rosenberg. Homogeneous vs. heterogeneous clustered sensor networks: A comparative study. In *In Proceedings of 2004 IEEE International Conference on Communications (ICC 2004*, pages 3646–3651, 2004.

[30] K. Pahlavan and A. H. Levesque. Wireless information networks. 1995.

[31] M. Segal. Fast algorithm for multicast and data gathering in wireless networks. *Inf. Process. Lett.*, 107(1):29–33, 2008.

[32] M. Segal and H. Shpungin. On construction of minimum energy k-fault resistant topologies. *Ad Hoc Networks*, 7(2):363–373, 2009.

[33] H. Shpungin and M. Segal. Near optimal multicriteria spanner constructions in wireless ad-hoc networks. *IEEE/ACM transaction on Networking*, 2009.

[34] J. Stanford and S. Tongngam. Approximation algorithm for maximum lifetime in wireless sensor networks with data aggregation. In *FOCS 1998*, pages 300–309, 1998.

[35] V. Stoumpos, A. Deligiannakis, Y. Kotidis, and A. Delis. Processing event-monitoring queries in sensor networks. In *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering*, pages 1436–1438. IEEE Computer Society, 2008.

[36] K. Sundaresan and S. Rangarajan. On exploiting diversity and spatial reuse in relay-enabled wireless networks. pages 13–22, 2008.

[37] A. Woo, T. Tong, and D. Culler. Taming the underlying challenges of reliable multihop routing in sensor networks. In *SenSys '03: Proceedings of the 1st international conference on Embedded networked sensor systems*, pages 14–27, New York, NY, USA, 2003. ACM.

[38] Y. Wu, S. Fahmy, and N. B. Shroff. On the construction of a maximum-lifetime data gathering tree in sensor networks: Np-completeness and approximation algorithm. In *INFOCOM*, pages 356–360, 2008.

[39] H. Zhang, H. Shen, and H. Tian. Reliable and real-time data gathering in multi-hop linear wireless sensor networks. In X. Cheng, W. Li, and T. Znati, editors, *Wireless Algorithms, Systems, and Applications*, volume 4138 of *Lecture Notes in Computer Science*, pages 151–162. Springer Berlin, 2006.

[40] L. Zhang and B.-H. Soong. Interference Distribution in the Nakagami Fading Wireless CDMA Ad Hoc Networks with Multi-Code Multi-Packet Transmission (MCMPT). In *30th Annual IEEE Conference on Local Computer Networks (LCN 2005), 15-17 November 2005, Sydney, Australia, Proceedings*. IEEE Computer Society, 2005.