Bounded-hop Energy-Efficient Broadcast in Low-Dimensional Metrics via Coresets *

Stefan Funke and Sören Laue

Max-Planck-Institut für Informatik,
Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany

Abstract. We consider the problem of assigning powers to nodes of a wireless network in the plane such that a message from a source node s reaches all other nodes within a bounded number k of transmissions and the total amount of assigned energy is minimized. By showing the existence of a *coreset* of size $O(\left(\frac{1}{\epsilon}\right)^{4k})$ we are able to $(1+\epsilon)$-approximate the bounded-hop broadcast problem in time *linear* in n which is a drastic improvement upon the previously best known algorithm.

While actual network deployments often are in a planar setting, the experienced metric for several reasons is typically not exactly of the Euclidean type, but in some sense 'close'. Our algorithm (and others) also work for non-Euclidean metrics provided they exhibit a certain similarity to the Euclidean metric which is known in the literature as *bounded doubling dimension*. We give a novel characterization of such metrics also pointing out other applications such as space-efficient routing schemes.

1 Introduction

Radio networks connecting a number of stations without additional infrastructure have recently gained considerable interest. Since the sites often have limited power supply, the energy consumption of communication is an important optimization criterion.

In the first part of the paper we consider the following problem: Given a set P of points (stations) in $\mathbb{R}^2$ and a distinguished source point $s \in P$ (sender) we want to assign distances/ranges $r : P \to \mathbb{R}_0^+$ to the elements in P such that the resulting communication graph contains a branching rooted at s spanning all elements in P and with depth at most k (an edge (p, q) is present in the communication graph iff $r(p) \geq |pq|$). Goal is to minimize the total assigned energy $\sum_{p \in P} r(p)^\delta$, where δ is the *distance-power gradient* and typically a constant between 2 and 6 ($\delta = 2$ reflects the exact energy requirement for free space communication, larger values are used as popular heuristic model for absorption effects). Such a branching corresponds to a *broadcast* operation from station s to all other nodes in the network with bounded latency. This is one of the most basic communication tasks in a wireless radio network.

* This work was supported by the Max Planck Center for Visual Computing and Communication (MPC-VCC) funded by the German Federal Ministry of Education and Research (FKZ 01IMC01).

In Section 2 of this paper we construct a (k, ϵ)-*coreset* of size $O(\left(\frac{1}{\epsilon}\right)^{4k})$ for a given instance of a bounded-hop broadcast problem, that is, we identify a small subset of the original problem instance for which the solution translates to an almost as good solution of the original problem. Interestingly, the size of this 'problem sketch' only depends on k and the desired approximation quality $(1 + \epsilon)$ but is independent of n. Hence we can approximate the bounded-hop broadcast problem – even using a brute force algorithm – in time *linear* in n and only doubly exponential in k (in contrast to the result in [1] which is *triply* exponential in k where it is also an exponent of n).

For analytical purposes it is very convenient to assume that all network nodes are placed in the Euclidean plane; unfortunately, in real-world wireless network deployments, especially if not in the open field, the experienced energy requirement to transmit does not exactly correspond to some power of the Euclidean distance between the respective nodes. Buildings, uneven terrain or interference might affect the transmission characteristics. Nevertheless there is typically still a strong correlation between geographic distance and required transmission power. An interesting question is now how to model analytically this correlation. One possible way is to assume that the required transmission energies are powers of the distance values in some metric space containing all the network nodes, and that this metric space has some resemblance to a low-dimensional Euclidean space. Resemblance to low-dimensional Euclidean spaces can be described by the so-called *doubling dimension* [5]. The *doubling dimension* of a metric space (X, d) is the least value α such that any ball in the metric with arbitrary radius R can be covered by at most 2^α balls of radius $R/2$. Note that for any $\alpha \in \mathbb{N}$, the Euclidean space $\mathbb{R}^\alpha$ has doubling dimension $\Theta(\alpha)$. In Section 3 we consider the doubling dimension a bit more in-depth and give a novel characterization of such metrics based on *hierarchical fat decompositions* (HFDs). We then show how the algorithm for energy-efficient broadcast presented in Section 2 as well as other algorithms in the wireless networking context can be adapted to metric spaces of bounded doubling dimension. Interestingly, metrics of bounded doubling dimension are not a tight characterization of all the metrics that allow for well-behaved HFDs, that is, there are metrics which are *not* of bounded doubling dimension, still our and many other algorithms run efficiently. As a side result we show how such HFDs directly lead to well-separated pair decompositions of linear-size (such WSPDs were also constructed in a randomized fashion in [7]). Finally, in Section 4 we examine metrics of bounded doubling dimension that arise as shortest-path metrics in unweighted graphs (e.g. unit-disk communication graphs). We show that for such metrics, an HFD can be computed in near-linear time, and the latter can be instrumented to derive a simple deterministic routing scheme that allows for $(1+\epsilon)$-stretch using routing tables of size $O((\frac{1}{\epsilon})^{O(\alpha)} \cdot \log^2 n)$ bits using a rather simple construction (compared to [3]).

Related Work

In [1] Ambühl et al. present an exact algorithm for solving the 2-hop broadcast problem with a running time of $O(n^7)$ as well as a polynomial-time approximation scheme for a fixed number of hops k and constant ϵ which has running time $O(n^{\mu})$ where $\mu = O((k^2/\epsilon)^{2^k})$, that is, their algorithm is *triply* exponential in the number of hops (and this dependence shows up in the exponent of n!). Both their algorithms are for the low-dimensional *Euclidean* case. Metrics of bounded doubling dimension have been studied for quite some time, amongst others Talwar in [9] provides algorithms for such metrics that $(1+\epsilon)$ approximate various optimization problems like TSP, k-median, and facility location. Furthermore he gives a construction of a well-separated pair decomposition for unweighted graphs of bounded doubling dimension α that has size $O(s^{\alpha} n \log n)$ (for doubling constant s). Based on that he provides compact representation schemes like approximate distance labels, a shortest path oracle, as well as a routing scheme which allows for $(1+\epsilon)$-paths using routing tables of size $O((\frac{\log n}{\epsilon})^{\alpha} \log^2 n)$. An improved routing scheme using routing tables of size $O((1/\epsilon)^{O(\alpha)} \log^2 n)$ bits was presented in [3] by Chan et al., but the construction is rather involved and based on a derandomization of the Lovasz Local Lemma. Har-Peled and Mendel in [7] gave a randomized construction for a WSPD of *linear* size which matches the optimal size for the Euclidean case from Callahan and Kosaraju in [2].

2 Bounded-hop Energy-Efficient Broadcast in $\mathbb{R}^2$

Given a set P of n nodes in the Euclidean plane, a *range assignment* for P is a function $r : P \to \mathbb{R}_{\geq 0}$. For a given range assignment r we define its overall power consumption as $\nu_r = \sum_{p \in P} (r(p))^{\delta}$. A range assignment r for a set P induces a *directed communication graph* $G_r = (P, E)$ such that for each pair $(p, q) \in P \times P$, the directed edge (p, q) belongs to E if and only if q is at distance at most $r(p)$ from p, i.e. $|pq| \leq r(p)$.

The *k-hop broadcast problem* is defined as follows. Given a particular source node s, G_r must contain a directed spanning tree rooted at source s to all other nodes $p \in P$ having depth at most k. W.l.o.g. we assume the largest Euclidean distance between the source node s and any other node $p \in P$ to be equal to 1. We say a range assignment r is *valid* if the induced communication graph G_r contains a directed spanning tree rooted at s with depth at most k; otherwise we call r *invalid*.

Definition 1. *Let P be a set of n points, $s \in P$ a designated source node. Consider another set S of points (not necessarily a subset of P). If for any valid range assignment $r : P \to \mathbb{R}_{\geq 0}$ there exist a valid range assignment $r' : S \to \mathbb{R}_{\geq 0}$ such that $\nu_{r'} \leq (1+\epsilon) \cdot \nu_r$ and for any valid range assignment $r' : S \to \mathbb{R}_{\geq 0}$ there exists a valid range assignment $r : P \to \mathbb{R}_{\geq 0}$ such that $\nu_r \leq (1+\epsilon) \cdot \nu_{r'}$ then S is called (k, ϵ)-**coreset** for (P, s).*

A (k, ϵ)-coreset for a problem instance (P, s) can hence be viewed as a problem sketch of the original problem. If we can show that a coreset of small size exists, solving the bounded-hop broadcast problem on this problem sketch immediately leads to an $(1+\epsilon)^2$-solution to the original problem.

This definition of a coreset differs slightly from the definition of a coreset defined in previous papers. For example, the term coreset has been defined for k-median [6] or minimum enclosing disk [8]. However, in the case of the bounded-hop broadcast problem we have to consider two more issues. The first is feasibility. While any solution to the coreset for the k-median problem is feasible wrt to the original problem this is not the case for every coreset solution for the bounded-hop broadcast problem. The second issue is monotonicity. For the problem of the smallest enclosing disk the optimal solution does not increase if we remove points from the input. We do not have this property here. An optimal solution can increase or decrease if we remove points.

Our coreset construction is heavily based on the insight that for any valid range assignment r there exists an almost equivalent (in terms of total cost) range assignment r' where all assigned ranges are either zero or rather 'large'. We formalize this in the following structure lemma:

Lemma 1 (Structure Lemma). *Let r be a valid range assignment for (P, s) of cost ν_r. For any $0 < \epsilon < 1$ there exists a valid range assignment r' with either $r'(p) = 0$ or $r'(p) \geq (1-\epsilon)\epsilon^{2k-2}$ and total cost $\nu_{r'} \leq \left(1 + \frac{\epsilon}{1-\epsilon}\right)^{\delta} \nu_r$.*

Proof: Let r be a valid range assignment. Consider a spanning tree rooted at s of depth at most k contained in the communication graph G_r. We call it the communication tree.

We will construct a valid range assignment r' from the given range assignment r. Initially, we set $r'(p) = r(p)$. After the first phase we will ensure $r'(s) \geq (1-\epsilon)\epsilon^{k-1}$ and after the second phase we will ensure $r'(p) \geq (1-\epsilon)\epsilon^{2k-2}$ for any node p.

The core idea to this construction is that if we have two nodes that are geometrically close to each other and one has a large power value $r(p)$ assigned to it and the other a rather small power value, we can safely increase the larger by a bit, remove the smaller one, and still have a valid power assignment. We apply this idea once in the opposite direction of the communication paths, i.e. towards the source node s (first phase) and once along the direction of the directed communication paths (second phase).

If $r(s) \geq (1-\epsilon)\epsilon^{k-1}$ we are done with the first phase. Otherwise, there exists a directed path of length at least 1 from source node s to some node p having at most k hops. Let the nodes on this path be labeled $p = p_0, p_1, \ldots, p_l = s$, $l \leq k$ as in Figure 1. Note that $r(p_0)$ does not contribute to the length of this path as it is the last node on the directed path. On this path pick the node with largest index j such that $r(p_j) \geq (1-\epsilon)\epsilon^{j-1}$. Such a node clearly exists as $\sum_{i=1}^{l} r(p_i) \geq 1$ and $\sum_{i=1}^{l}(1-\epsilon)\epsilon^{i-1} < 1$. Setting $r'(s) = r(p_j)\left(1 + \frac{\epsilon}{1-\epsilon}\right)$ and $r'(p_i) = 0$ for $i = j \ldots l-1$ increases the cost $\nu_{r'}$ only slightly but still ensures

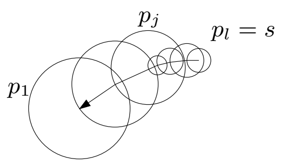

Fig. 1. Original range assignment before the first phase

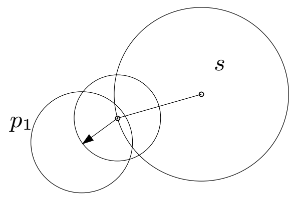

Fig. 2. Range assignment after the first phase

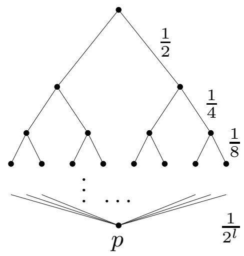

Fig. 3. A metric with unbounded doubling dimension but with bounded degree HFD

a valid range assignment because

$$r'(s) = r(p_j)\left(1 + \frac{\epsilon}{1-\epsilon}\right) \geq r(p_j) + \epsilon^j > r(p_j) + \sum_{i=j+1}^{l} (1-\epsilon)\epsilon^{i-1} \tag{1}$$

$$> r(p_j) + \sum_{i=j+1}^{l} r(p_i), \tag{2}$$

i.e. we increased $r'(s)$ such that all nodes that could be reached by nodes $p_j, p_{j+1}, \ldots, p_{l-1}$ can now be reached directly by s.

In the second phase we can use an analogous argument starting from source node s. We assign each node p in the communication tree a level according to the number of hops to the source node s, where the source node s has level 0 and the leaves of the tree have level at most k. We distinguish two cases. In the first case $r'(s) = r(s)$, i.e. the value of the starting node s has not been increased. The other case occurs when it has been increased, i.e. $r'(s) > r(s)$.

Let us look at the first case. Consider all maximal paths $\{t_j\}$ in the communication tree starting from node s where all nodes have $r(p) < (1-\epsilon)\epsilon^{k-1+i}$ if node p is on level i. We can set $r'(s) = r(s)(1 + \frac{\epsilon}{1-\epsilon})$ and $r'(p) = 0$ for all $p \in t_i$. Hence, we again maintain a valid range assignment and the next nodes p along the paths of the communication tree satisfy $r(p) \geq (1-\epsilon)\epsilon^{k-1+i}$ if node p is on level i. Applying the same reasoning iteratively to these nodes we finally have that for all nodes p either $r'(p) = 0$ or $r'(p) \geq (1-\epsilon)\epsilon^{k-1+i}$ for a node p on level i. Note that for nodes p on level k we can set $r'(p) = 0$. Hence, we have a valid range assignment r' with $r'(p) \geq (1-\epsilon)\epsilon^{2k-2}$.

Let us now consider the second case, when $r'(s) > r(s)$, i.e. the value of s has been increased in the first phase of the construction. Here we increased $r'(s)$ already in the first phase to at least $(1-\epsilon)\epsilon^{k-2}\left(1 + \frac{\epsilon}{1-\epsilon}\right) = \epsilon^{k-2}$. Hence, we can continue as in the first case without increasing $r'(s)$ anymore, because $\epsilon^{k-2} > \sum_{i=0}^{k}(1-\epsilon)\epsilon^{k-1+i}$ for $\epsilon < 1$.

The cost of the valid range assignment r' satisfies

$$\nu_{r'} = \sum_{p\in P} (r'(p))^{\delta} \leq \sum_{p\in P} \left(r(p) \left(1 + \frac{\epsilon}{1-\epsilon} \right) \right)^{\delta} = \left(1 + \frac{\epsilon}{1-\epsilon} \right)^{\delta} \nu_r \qquad (3)$$

■

Using the preceding Lemma it is now easy to come up with a small coreset by using a grid of width roughly an ϵ-fraction of the minimum non-zero range assigned in r'.

Lemma 2. *For any k-hop broadcast instance there exists a $(k, (\delta+2)\epsilon)$-coreset of size $O((\frac{1}{\epsilon})^{4k})$.*

Proof: We will only sketch the main idea here. We place a grid of width $\Delta = \frac{1}{\sqrt{2}}\epsilon \cdot r_{min}$ on the plane, where $r_{min} = (1-\epsilon)\epsilon^{2k-2}$. Notice, that the grid has to cover an area of radius 1 around the source only because the furthest distance from node s to any other node is 1. Hence its size is $O((\frac{1}{\epsilon})^{4k})$ for small ϵ. Now assign each point in P to its closest grid point. Let the coreset S be the set of grid points that had at least one point from P snapped to it. Applying the Structure Lemma 1 induces a relative error of $\left(1 + \frac{\epsilon}{1-\epsilon}\right)^{\delta}$. Since the grid induces an error of $(1+\epsilon)$ the total relative error is bounded by $(1 + (\delta+2)\epsilon)$. ■

Unfortunately we are not aware of any efficient algorithm for computing even just a constant approximation to the bounded-hop broadcast problem. But since we were able to reduce the problem size to a constant independent of n, we can also employ a brute-force strategy to compute an optimal solution for the reduced problem (S, s), which in turn translates to an $(1+(\delta+2)\epsilon)^2$-approximate solution to the original problem since the reduced problem (S, s) is a $(k, (\delta+2)\epsilon)$-coreset.

When looking for a optimal, energy-minimal solution for S, it is obvious that each node needs to consider only $|S|$ different ranges. Hence, naively there are at most $|S|^{|S|}$ different range assignments to consider at all. We enumerate all these assignments and for each of them we check whether the induced communication graph contains a directed spanning tree of depth at most k rooted at the grid point corresponding to the original root node s, that is whether the respective range assignment is valid; this can be done in time $|S|^2$. Of all the valid range assignments we return the one of minimal cost.

Assuming the floor function a $(k, (\delta+2)\epsilon)$-coreset S for an instance of the k-hop broadcast problem for a set of n radio nodes in the plane can be constructed in linear time. Hence we obtain the following corollary:

Corollary 1. *A $(1+(\delta+2)\epsilon)^2$-approximate solution to the k-hop energy-minimal broadcast problem on n points in the plane can be computed in time $O(n + |S|^{|S|}) = O\left(n + (\frac{1}{\epsilon})^{4k(\frac{1}{\epsilon})^{4k}}\right)$.*

A simple observation allows us to improve the running time slightly. Since eventually we are only interested in an approximate solution to the problem, we are also happy with only approximating the optimum solution for the coreset S. Such an approximation for S can be found more efficiently by not considering all possible at most $|S|$ ranges for each grid point. Instead we consider as admissible ranges only 0 and $r_{\min} \cdot (1+\epsilon)^i$ for $i \geq 0$. That is, the number of different ranges a node can attain is at most $1 + \log_{1+\epsilon} r_{min}^{-1} \leq \frac{4k}{\epsilon} \cdot \log \frac{1}{\epsilon}$ for $\epsilon \leq 1$. This comes at a cost of a $(1+\epsilon)$ factor by which each individual assigned range might exceed the optimum. The running time of the algorithm improves, though, which leads to our main result in this section:

Corollary 2. *A $(1+(\delta+2)\epsilon)^3$-approximate solution to the k-hop energy-minimal broadcast problem on n points in the plane can be computed in time* $O\left(n + \left(\frac{4k}{\epsilon} \cdot \log \frac{1}{\epsilon}\right)^{|S|}\right) = O\left(n + \left(\frac{4k}{\epsilon}\right)^{(\frac{1}{\epsilon})^{4k}}\right)$.

A $(1+\psi)$-approximate solution can be obtained by choosing $\epsilon = \theta(\psi/\delta)$.

3 Properties of Low-Dimensional Metrics

As mentioned in the introduction, the theoretical analysis of algorithms typically requires some simplifying assumptions on the problem setting. In case of wireless networking, a very common assumption is that all the network nodes are in the Euclidean plane, distances are the natural Euclidean distances, and the required transmission energy is some power of the Euclidean distance. This might be true for network deployments in the open field, but as soon as there are buildings, uneven terrain or interference, the effective required transmission power might be far higher. Still, it is true that there is a strong correlation between geographic/Euclidean distance and required transmission power. So how could we define the problem using less demanding assumptions but still be able to analytically prove properties of the algorithms and protocols of interest? One possible way is to assume that the required transmission energies are powers of distance values in some metric space on the network nodes, and that this metric space has some resemblance to a low-dimensional Euclidean space. "Resemblance to a low-dimensional Euclidean space" could be equivalent to the existence of a mapping into low-dimensional Euclidean space which more or less preserves distances (low distortion embeddings). Another means to capture similarity to low-dimensional Euclidean spaces is the so-called *doubling dimension*. The *doubling dimension* of a metric space (X, d) is the least value α such that any ball in the metric with arbitrary radius R can be covered by at most 2^α balls of radius $R/2$. Note that for any $\alpha \in \mathbb{N}$, the Euclidean space $\mathbb{R}^\alpha$ has doubling dimension $\Theta(\alpha)$. In the following we show that a metric of bounded doubling dimension exhibits not only this Euclidean-like covering property but also a respective *packing* property.

3.1 Metrics of Bounded Doubling Dimension

The fact that every ball can be covered by at most a constant number of balls of half the radius (covering property) induces the fact, that not too many balls of sufficiently large radius can be placed inside a larger ball (packing property). The following lemma states this fact precisely. (The same observation was made in Section 2 of [7] in the context of net-trees but was not explicitly stated in this general form.)

Lemma 3 (Packing Lemma). *Given a metric (X, d) with doubling constant k, i.e. every ball can be covered by at most k balls of half the radius, then, at most k pairwise disjoint balls of radius $r/2 + \epsilon$, for $\epsilon > 0$ can be placed inside a ball of radius r.*

Proof: Consider a ball B of radius r. Place a set $S = \{B_1, B_2, \ldots, B_l\}$ of pairwise disjoint balls each having radius $r/2+\epsilon$ inside B. Let $C = \{b_1, b_2, \ldots, b_k\}$ be a set of balls of radius $r/2$ that cover the ball B. The distance between two centers of balls from S is at least $r + 2\epsilon > r$ as they are pairwise disjoint. Hence, every ball $b_i \in C$ can cover at most one center of a ball $B_j \in S$. Since every ball from the set S is covered and especially its center, we have $|S| \leq |C| = k$. ■

The same generalizes to arbitrary radii. If a ball B of radius R can be covered by at most k balls of radius r then there can be at most k pairwise disjoint balls of radius $r + \epsilon$ for $\epsilon > 0$ placed inside B. We will make use of this packing property at various places later.

3.2 Hierarchical Fat Decompositions (HFD)

Given an arbitrary metric (X, d), a *decomposition* is a partition of X into clusters $\{C_i\}$. A *hierarchical decomposition* is a sequence of decompositions $P_l, P_{l-1}, \ldots, P_0$, where each cluster in P_i is the union of clusters from P_{i-1}, $P_l = X$, and $P_0 = \{\{x\} | x \in X\}$, i.e. P_l is the single cluster containing X and every point forms one separate cluster in P_0.[1] We refer to clusters of P_i as clusters at level i. A hierarchical decomposition where each cluster of the same level i is contained in a ball of radius r_i, contains a ball of radius $\alpha \cdot r_i$, and $r_{i-1} \leq \beta \cdot r_i$ for constants α and $\beta < 1$ is called a *hierarchical fat decomposition* (HFD). Thus, in an HFD clusters are fat and the size of the clusters from different levels form a geometric sequence. We call a set *fat* if the ratio between an inscribed ball and a surrounding ball is bounded by a constant.

We will show how to construct an HFD for an arbitrary metric (X, d). Without loss of generality we assume $min_{p,q \in X} d(p, q) = 1$. We call $\Phi = max_{p,q \in X} d(p, q)$ the *spread* of X. We construct the HFD bottom-up. Let L_i be a set of points which we call landmarks of level i. With each landmark we associate a cluster $C_i(l) \subseteq X$.

On the lowest level we have $L_o = X$ and $C_0(l) = \{l\}$, i.e. each point forms a separate cluster. Obviously, each cluster is contained in a ball of radius 1 and contains a ball of radius $\frac{1}{2}$. Starting from the lowest level we construct the next level recursively as follows. For level i we compute a 4^i-independent maximal set (i.e. a maximal set with respect to insertion with the pairwise distance of at least 4^i) of landmarks L_i from the set L_{i-1} of landmarks from one level

[1] This is also known as a laminar set system as used frequently in the literature.

below. Hence, the distance between any two landmarks of level i is at least 4^i. We compute the Voronoi diagram VD of this set L_i and call the Voronoi cell of l $VC_i(l)$. The union of all clusters of landmarks from level $i-1$ that fall in the region $VC_i(l)$ form the new cluster that we associate with landmark l, i.e. $C_i(l) = \bigcup_{p \in VC_i(l)} C_{i-1}(p)$. Obviously, each Voronoi cell contains a ball of radius $4^i/2$ and is contained in a ball of radius 4^i, since the set of landmarks L_i form a 4^i maximal independent set. Hence, each cluster on level i is contained in a ball of radius $\sum_{j=0}^{i} 4^j \leq 4^{i+1}/3$ and each cluster contains a ball of radius $4^i/2 - \sum_{j=0}^{i-1} 4^i \geq 4^i/6$. Thus, we have constructed an HFD.

3.3 A Characterization of Metrics of Bounded Doubling Dimension

We say an HFD has degree d if the tree induced by the hierarchy has maximal degree d. The following theorem gives a characterization of metrics with bounded doubling dimension in terms of such HFDs.

Theorem 1. *A metric (X, d) has bounded doubling dimension if and only if all hierarchical fat decompositions of (X, d) have bounded degree.*

Proof: First, suppose metric (X, d) has bounded doubling dimension. Fix an arbitrary HFD for (X, d) and pick a cluster C. Since C is fat, it is contained in a ball of radius r_1 and it is the union of fat clusters $\{C_1, C_2, \ldots, C_l\}$. Each of them contains a ball of radius r_2. The ratio of the two radii r_1 and r_2 is bounded by a constant due to the definition of an HFD. Then, by the Packing Lemma 3 cluster C cannot contain more than a constant number of clusters from the level below. Hence, each HFD has bounded degree.

On the other hand, suppose (X, d) has no bounded degree. Then there exists a ball $B(x, r) = \{y | d(x, y) \leq r\}$ that cannot be covered by a constant number of balls of half the radius r. We can construct an HFD, which has no bounded degree as follows. Consider an HFD constructed as in Section 3.2, where the set of landmarks always contains the point x. Consider the minimal cluster C that contains ball $B(x, r)$ and consider the set of children clusters $\{C_1, C_2, \ldots, C_l\}$ of C that are all contained in a ball of radius $r/2$. Due to the definition of an HFD the difference in the levels of these clusters is bounded by a constant. Since, the number of children clusters is not bounded, the HFD cannot have bounded degree. ■

There are metrics however, that admit an HFD with bounded degree but do not have bounded doubling dimension. The following metric is such an example. Consider the complete binary tree of depth l and each edge from level $i-1$ to level i having weight $\frac{1}{2^i}$ as in Figure 3. Let p be a node which is connected to all leaves with edge weights $\frac{1}{2^l}$. The shortest path metric induced by this graph does not have a bounded doubling dimension but admits an HFD with bounded degree. We can place 2^l disjoint balls of radius $\frac{1}{2^{l+1}}$, each having a leaf as its center, inside a ball of radius $\frac{1}{2^l}$ with center p. Hence, the metric cannot have bounded doubling dimension for arbitrary large l (Packing Lemma). On the other hand, it is easy to see that the metric has an HFD of degree 2.

An HFD with bounded degree immediately implies a well-separated pair decomposition (WSPD) of linear size in the number of input points. We just sketch the main idea here.

The construction follows closely the lines of [2]. If we replace in their construction the *fair split tree* by our hierarchical fat decomposition, we get the same bounds, apart from constant factors. All we need to show is that if a ball B of radius r is intersected by the surrounding balls of a set of clusters $S = \{C_1, C_2, \ldots, C_l\}$ with $C_j \cap C_j = \emptyset$ for $i \neq j$ and the parent of each cluster C_i has a surrounding ball of radius larger than r/c for a constant c, then the set S can only contain a constant number of clusters. But this is certainly true. The packing lemma 3 assures that there are just a constant number of clusters whose surrounding balls intersect a large ball B whose radius is larger by a constant. And as the HFD has bounded degree, these clusters have constant number of children clusters $S = \{C_1, C_2, \ldots, C_l\}$ all together. If we eliminate all clusters in the HFD that just have one children cluster we get that the number of well-separated pairs is linear in the number of input points and depends only on the constant c and the doubling dimension.

3.4 Optimizing Energy-Efficiency in Low-Dimensional Metrics

In the following we will briefly sketch how the algorithm presented in Section 2 can also be applied for metrics of bounded doubling dimension. Furthermore we show how an old result ([4]) can also be partly adapted from the Euclidean setting.

Energy-Efficient k-hop Broadcast The algorithm presented in Section 2 for broadcasting in the plane can be generalized to metrics with bounded doubling dimension. Obviously, the Structure Lemma 1 still holds since the triangle inequality holds. Now, instead of placing a planar grid, we construct an HFD for the nodes as in Section 3.2. The level of the decomposition where each cluster is contained in a ball of radius $r = \Delta/2$ replaces the grid in the approximation algorithm. As the metric has bounded doubling dimension, the HFD has bounded degree. Hence, there is just a constant number of clusters in the decomposition of this level. We can solve this instance in the same way as for the planar case.

Energy-Efficient k-hop paths In [4] the authors considered the problem of computing an $(1+\epsilon)$ energy-optimal path between a nodes s and t in a network in $\mathbb{R}^2$ which uses at most k hops. Again, as in Section 2, the assumption was that the required energy to transmit a message over Euclidean distance d is d^δ, for $\delta \geq 2$. Using a rather simple construction where the neighborhood of the query pair s and t was covered using a constant number of grid cells (depending only on k, δ, ϵ) such queries could be answered with a $(1+\epsilon)$ guarantee in $O(\log n)$ time. Similarly to the bounded-hop broadcast, we can replace this grid by a respective level in a HFD. For bounded doubling dimension we then know that there are only a constant number of relevant grid cells and the algorithm can be implemented as in the Euclidean case. In [4] the construction was further refined by using a WSPD to actually precompute a linear number of k-hop paths which then could be accessed in $O(1)$ time for a query (independent of k, δ, ϵ). Generalizing this refinement is the focus of current research.

4 Computing HFDs in Shortest-Path Metrics

In wireless sensor networks, the employed network nodes are typically low-capability devices with simple computing and networking units. In particular, most of these devices do not have the ability to adjust the transmission power but always send within a fixed range. The graph representing the pairs of nodes that can communicate with each other is then a so-called *unit-disk graph* (UDG), where two nodes can exchange messages directly iff they are at distance of most 1. Typically UDGs are considered in the Euclidean setting, but they can be looked at in any metric space. Due to the fixed transmission power, saving energy by varying the latter is not possible. Still, indirectly, energy can be saved by for example better routing schemes which yield shorter (i.e. fewer hops) paths. In the following we briefly discuss how HFDs can be used to provide such efficient routing schemes. We first show how in case of unweighted graphs like UDGs, HFDs can be efficiently computed and then sketch how the structure of the HFDs can be exploited to allow for routing schemes with near-optimal path lengths using small routing tables at each node.

4.1 A Near-Linear Time Algorithm

Consider an unweighted graph $G = (V, E)$. All shortest paths define a shortest-path metric on the set of vertices. If the metric has bounded doubling dimension we can construct an HFD with bounded degree efficiently by employing the generic approach described in Section 3.2. At level i we need to construct an 4^i-independent maximal set of nodes L_i, the landmarks. This can be done greedily using a modified breadth-first search algorithm on the original graph G. At the same time we can compute the corresponding Voronoi diagram. We pick an arbitrary node n_1 and add it to the set L_i. In a breadth-first search we successively compute the set of nodes that have distance 1, 2, ... until we computed the set of nodes at distance 4^i. We mark each visited node as part of the Voronoi cell of node n_1 and store its distance to n_1. From the set of nodes at distance 4^i we pick a node n_2 and add it to L_i. Starting from node n_2 we again compute the set of nodes that have distance 1, 2, ... to the node n_2. Similarly, if a node is not assigned to a Voronoi cell, we assign it to n_2. If it has been assigned already to some other node but the distance to the other landmark is larger than to the current node n_2, we reassign it to the current node. We do this until no new landmark can be found and all nodes are assigned to its Voronoi cell.

We might visit a node or an edge several times, but as the metric has bounded doubling dimension, this happens only a constant number of times. Thus, the running time is $O(m + n)$ for one level and $O((m + n) \log n)$ for the whole construction of the HFD as there are $O(\log n)$ levels.

4.2 Hierarchical Routing in Doubling Metrics

The HFD constructed above implicitly induces a hierarchical naming scheme for all nodes of the network by building IP-type addresses which reflect in which child cluster of each level a node v is contained (remember that there are always only a constant number of children of each cluster). For example if v is contained in the top-most cluster 4, in the 2nd child of that top-most cluster and in the 5th child of that child, its name would be 4.2.5. Clusters can be named accordingly and will be prefixes of the node names. We now install routing tables at each node which allow for almost-shortest path routing in the network: For every cluster $\mathcal{C}$ with diameter D we store at all nodes in the network which have distance at most $O(D/\epsilon)$ from $\mathcal{C}$ a distance value (associated with the respective address of the cluster and a pointer to the predecessor on the shortest path to the cluster) to the boundary of $\mathcal{C}$ in the node's routing table. Now, when a message needs to be routed to a target node t and is currently at node p, p inspects its routing table and looks for an entry which is a as large as possible prefix of the target address. p then forwards the message to the adjacent neighbor which is associated with this routing table entry. A simple calculation shows that this yields paths which are at most a $(1+\epsilon)$ factor longer than the optimal shortest path For the size of the routing table first consider an arbitrary node v and clusters of diameter at most D. Clearly there are at most $O((1/\epsilon)^{O(\alpha)})$ many such clusters which have distance less than $O(D/\epsilon)$ from v and have hence created a routing table entry at v. Overall there are only $\log n$ levels and each routing table entry has size $O(\log n)$ (since the maximum distance is n). Hence the overall size of the routing table of one node is $O((1/\epsilon)^{O(\alpha)} \log^2 n)$.

References

1. C. Ambühl, A. E. F. Clementi, M. Di Ianni, N. Lev-Tov, A. Monti, D. Peleg, G. Rossi, and R. Silvestri. Efficient algorithms for low-energy bounded-hop broadcast in ad-hoc wireless networks. In *STACS*, pages 418–427, 2004.
2. Paul B. Callahan and S. Rao Kosaraju. Algorithms for dynamic closest pair and n-body potential fields. In *SODA*, 1995.
3. Hubert T.-H. Chan, Anupam Gupta, Bruce M. Maggs, and Shuheng Zhou. On hierarchical routing in doubling metrics. In *SODA*, pages 762–771, 2005.
4. S. Funke, D. Matijevic, and P. Sanders. Approximating energy efficient paths in wireless multi-hop networks. In *ESA*, pages 230–241, 2003.
5. A. Gupta, R. Krauthgamer, and J. R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *FOCS*, 2003.
6. S. Har-Peled and S. Mazumdar. Coresets for k-means and k-median clustering and their applications. In *STOC*, pages 291–300, 2004.
7. S. Har-Peled and M. Mendel. Fast construction of nets in low dimensional metrics, and their applications. *SIAM Journal on Computing*, 35(5):1148–1184, 2006.
8. P. Kumar, J. S. B. Mitchell, and E. A. Yildirim. Approximate minimum enclosing balls in high dimensions using core-sets. *J. Exp. Algorithmics*, 8, 2003.
9. Kunal Talwar. Bypassing the embedding: algorithms for low dimensional metrics. In *STOC*, pages 281–290, 2004.