University of Twente (26 October 2007)

AN AVERAGE CASE ANALYSIS OF THE MINIMUM SPANNING TREE HEURISTIC FOR THE RANGE ASSIGNMENT PROBLEM

RICHARD J. BOUCHERIE,* *University of Twente*

MAURITS DE GRAAF,** *Thales Land & Joint Systems*

Abstract

We present an average case analysis of the minimum spanning tree heuristic for the range assignment problem on a graph with power weighted edges. It is well-known that the worst-case approximation ratio of this heuristic is 2. Our analysis yields the following results: (1) In the one dimensional case ($d = 1$), where the weights of the edges are 1 with probability p and 0 otherwise, the average-case approximation ratio is bounded from above by $2-p$. (2) When $d = 1$ and the distance between neighboring vertices is drawn from a uniform $[0, 1]$-distribution, the average approximation ratio is bounded from above by $2-2^{-\alpha}$ where α denotes the distance power gradient. (3) In Euclidean 2-dimensional space, with distance power gradient $\alpha = 2$, the average performance ratio is bounded from above by $1 + \log 2$.

Keywords: average case analysis; range assignment; power assignment; ad-hoc networks; analysis of algorithms; approximation algorithms; point processes
AMS 2000 Subject Classification: Primary 68W40
Secondary 68W25

1. Introduction

Ad hoc wireless networks have received significant attention in recent years due to their potential applications in battlefield, emergency disaster relief, and other scenarios (see, for example [13], [18], and [19]). In an ad hoc wireless network, a communications session is achieved either through single-hop transmission, if the recipient is within

* Postal address: University of Twente, P.O. Box 217, 7500 AE Enschede, Netherlands
** Postal address: Thales, P.O. Box 88, 1270 AB Huizen, Netherlands, e-mail: maurits.degraaf@nl.thalesgroup.com

the transmission range of the source, or by relaying through intermediate nodes. We assume an idealized propagation model, where omnidirectional antennas are used by all nodes to transmit and receive signals. Thus, a transmission made by a node can be received by all nodes within its transmission range.

The topology of a multihop wireless network is the set of communication links between node pairs. The topology depends on uncontrollable factors such as node mobility, weather, interference, noise as well as on controllable parameters such as transmit power. We assume that for the purpose of energy conservation, each node can adjust its transmit power.

In this paper we analyze an algorithm to control the topology of the network by changing the transmit powers of the nodes. Two extreme approaches exist: if the transmit powers assigned to the nodes are too low, the resulting topology may be too sparse and the network may be partitioned. On the other extreme, if the transmit powers assigned to the nodes are too high, the limited spatial reuse reduces network capacity and nodes run out of energy quickly.

The goal of the Connected Minimum Power Assignment (CMPA-) problem is to assign transmission powers to the transceivers so that the resulting network is connected and the sum of transmit powers assigned to the transceivers is minimized (see e.g. [13]).

This problem is, in general, NP-hard (for some special cases there are polynomial solutions). A well-known approximation exists: the Minimum Spanning Tree (MST-) heuristic. This heuristic is known to have a worst-case approximation ratio of 2 (see e.g. [12]). This paper presents an average case analysis of the MST-heuristic for the range assignment problem.

1.1. Notation and previous work

Formally, for a set of points V representing the nodes in a network, a power assignment is a function $p : V \to \mathbb{R}$. Following the notation of [13], for each ordered pair (u, v) of transceivers, there is a transmit power threshold, denoted by $c(u, v)$, with the following meaning: a signal transmitted by the transceiver u can be received by v only when the transmit power is at least $c(u, v)$. In our approach, we assume that for each pair of points the transmit power threshold values $c(u, v)$ are known, and that these values are symmetric, i.e., $c(u,v) = c(v,u)$ for all pairs $\{u,v\} \in V$. This way, a power assignment p directly defines the undirected graph $G_p = (V, E_p)$ where an edge $e = \{u,v\} \in E_p$ if and only if $p(u) \geq c(u,v)$ and $p(v) \geq c(u,v)$.

This paper is concerned with the CMPA problem: given a graph $G = (V, E)$ with edge costs $c : E \to \mathbb{R}$, one asks for a power assignment $p : V \to \mathbb{R}$ such that G_p is connected and the total power $\sum_{v \in V} p(v)$ is minimal.

Often, each $v \in V$ has a given location in $\mathbb{R}^d$. In such cases a power attenuation model is assumed, where the signal power falls with the radius r as $\frac{1}{r^\alpha}$ where $\alpha \in \mathbb{R}$, called the *distance-power gradient*. According to [16], α depends on the wireless environment, and realistic values range from 1 to more than 6. This implies that the power required to support a link between two nodes separated by a distance r is r^α. In this case, the power assignment problem is also known as the *range assignment* problem, as assigning a power p_v to a node v corresponds to assigning a range r_v to a node v. The range assignment problem asks for minimization of the sum $\sum_{v \in V} r_v^\alpha$. Note that the power assignment problem is more general than the range assignment problem, as the weights are not necessarily based on a distance function. Note also that for this idealized setting, α is assumed to be constant for the whole problem (in reality, typically different values for α occur in different parts of the network). While the main results of this paper relate to the range assignment problem, intermediate results are derived for the power assignment problem.

The range assignment problem is NP-hard in all dimensions $d \geq 2$ for all values of the distance-power gradient α. In [9] and [3] the complexity of various other variants of this problem is analyzed.

The first NP-hardness result for the 3 dimensional range assignment problem was given by Kirousis et al. [12]. Clementi et al. showed NP hardness in 2 dimensions [5]. Therefore, polynomial time approximation algorithms are studied. The earliest approximation algorithm is the Minimum Spanning Tree (MST)-algorithm (see [9], [4]).

MST-Algorithm

1. Given a graph (V, E, c) compute a minimum spanning tree T using c as edge costs.

2. For each node $v \in V$: $p(v) = \max\{c(e)|e \text{ incident to } v \text{ in } T\}$

Other approximation algorithms are studied in [2], where a polynomial time approximation scheme with performance ratio approaching 5/3 as well as a more practical approximation algorithm with approximation factor 11/6 are given.

From now on, we write T_n for a minimum spanning tree of a graph on n-vertices. In addition, P_{T_n} denotes the power assignment *corresponding to* T_n, that is: for each $v \in V$: $p_T(v) = \max\{c(e)\|e \in Tande \text{ incident to } v\}$. When it is clear from the context which T_n is meant, we simply write P_n instead of P_{T_n}. We define $W(T_n)$ to be the total weight of the minimum spanning tree, and $W(P_n)$ for the total weight of the corresponding power assignment. It is well established (see e.g. [2], [4]) that

$$W(T_n) \leq W(P) \leq W(P_n) \leq 2W(T_n) < 2W(P_n) \tag{1}$$

where $W(P)$ denotes the weight of the optimal power assignment P. In [2] it is shown that in this statement the factor 2 cannot be replaced by a lower value.

While such a high factor of 2 might discourage use of this algorithm in practice, many papers present numerical results indicating that the MST algorithm is often rather close to the optimal solution.

The contribution of this paper consists of an analysis of the average case behavior of the function $W(P_n)/W(T_n)$ for $n \to \infty$ which provides an upper bound to the performance ratio $W(P_n)/W(P)$. To our knowledge, the average case behavior of the MST algorithm has never been analyzed. A probabilistic analysis of the range assignment problems has been performed in [20] focusing on upper- and lower-bounds for connectedness in case all nodes have the same transmission power.

The paper is structured as follows. In Section 2 we provide an observation bounding the weight of the power assignment in terms of the highest cost edges of the MST. Section 3 analyzes the 1-dimensional case for edge weights $\in \{0, 1\}$ and for uniformly distributed edge weights on $[0, 1]$. Section 4 presents results for the d-dimensional case where $d \geq 2$. Finally, Section 5 presents conclusions and directions for further research.

2. Observation

Let $G = (V, E)$ be any graph where to each edge $e \in E$ a cost $c(e) \in \mathbb{R}$ is assigned. Consider a minimum spanning tree T_n, with edges $e_1, \ldots, e_{n-1}$, where $c(e_1) \leq c(e_2) \leq \ldots \leq c(e_{n-1})$. We say that an edge $e \in E$ incident to v *covers* a vertex v, if $e \in T_n$, so $e = e_i$ for some $i \in 1, \ldots, n-1$, $c(e_i) = \max\{c(e_j)|\ e_j \text{ incident to } v\}$, and e_i has the maximal index i among maximum-weight edges of equal cost incident to v. Let $f(e)$ denote the number of nodes covered by $e \in T_n$, called the *covering number* of $e \in E$. Note that $f(e) \in \{0, 1, 2\}$. We immediately see that $\sum_{e \in E} f(e) = n$ as each vertex is covered exactly once. Moreover, $W(P_n) = \sum_{e \in T_n} f(e)c(e)$,

The main observation that enables us to bound the average case behavior of the MST algorithm for the range assignment problem is described in Lemma 1 which strengthens (1).

Lemma 1. *Let the edges $e_1, \ldots, e_{n-1}$ of a minimum spanning tree T_n be sorted such that $c(e_1) \leq c(e_2) \leq \ldots \leq c(e_{n-1})$. Let P_{T_n} denote the power assignment corresponding to T_n. Then*

$$W(P_{T_n}) = \sum_{i=1}^{n-1} f(e_i)c(e_i) \quad \leq \quad c_{\lfloor n/2 \rfloor} + 2 \sum_{i=\lceil n/2 \rceil}^{n-1} c(e_i) \ \textit{if } n \textit{ is odd} \tag{2}$$

$$\leq \quad 2 \sum_{i=n/2}^{n-1} c(e_i) \ \textit{if } n \textit{ is even} \tag{3}$$

and

$$W(P_{T_n}) = \sum_{i=1}^{n-1} f(e_i)c(e_i) \geq 2c(e_{n-1}) + \sum_{i=1}^{n-2} c(e_i) \tag{4}$$

Proof. From the fact that the maximum of the expression $W(P_{T_n}) = \sum_{i=1}^{n-1} f(e_i)c(e_i)$ is attained when f takes its maximum value at the highest weights (2) and (3) follow. Equation (4) can be inferred by induction as follows. For $n = 2$ (4) is clearly true. In T_n there are at least two edges incident to a vertex v with degree 1 (in T_n). Now choose e to be such an edge, so that $e \neq e_{n-1}$, and let k be the index so that $e = e_k$. It follows that $f(e_k) \geq 1$. Let G' be obtained from G by removing vertex v and edge e_k, and consider T_{n-1} obtained from T_n by removing e_k and v. By the choice of e_k, T_{n-1} is a spanning tree of G'. By induction hypothesis $W(P_{T_{n-1}}) \geq 2c(e_{n-1}) + \sum_{i=1, i\neq k}^{n-2} c(e_i)$.

Moreover we have $W(P_{T_n}) \geq W(P_{T_{n-1}}) + c(e_k)$. This completes the proof. (Note that strict inequality can only hold if $f(e) = 2$.)

Example 1. (*Tight bounds for inequalities (2), (3) and (4).*) Let $n = 2m+1$, suppose $G = (V, E)$ is a path $e_1, \ldots, e_{2m}$ so that $c(e_j) = 1$ if j is odd, and $c(e_j) = \epsilon < 1$ if j is even. G has only one spanning tree $T_n = G$. Sorting the edges according to increasing costs we first obtain m edges of cost ϵ, followed by m edges of cost 1. Moreover, $W(T_n) = m + m\epsilon$. Clearly, all edges with an odd index have covering number 2, and there is only one edge (being e_{2m}) with covering number 1, incident to the last vertex. So $W(P_n) = 2m + \epsilon$, which exactly corresponds to (2).

Let $n = 2m$, suppose G=(V,E) is a path $e_1, \ldots, e_{2m-1}$ so that $c(e_j) = 1$ if j is odd, and $c(e_j) = \epsilon < 1$ if j is even. Again, sorting the edges according to increasing costs we first obtain $m-1$ edges of cost ϵ, followed by m edges of cost 1. Moreover, $W(T_n) = m + (m-1)\epsilon$. Clearly, $W(P_n) = 2m$, which exactly corresponds to (3).

An example for equality in (4) is obtained by considering a graph $G = (V, E)$ where all costs $c(e)$ are 1. In this case $W(T_n) = n - 1$ and $W(P_n) = n$.

Note that from Lemma 1 it directly follows that for all n we have:

$$W(P_n) \leq 2 \sum_{i=\lfloor n/2 \rfloor}^{n-1} c(e_i). \tag{5}$$

This simplified inequality is used in the rest of this paper.

3. 1 dimension: the Spanning Tree is a path

3.1. 0,1 - weights

In the first part of the section we discuss the situation where $G = (V, E)$, all elements of V are on a line, each edge connects neighboring vertices (so $\|E\| = n - 1$), and the cost $c(e)$ of edge $e \in E$ is 1 with probability p and 0 with probability $1-p$. Generalizing the previous notation to random variables, $W(T_n)$ and $W(P_n)$ are now considered as random variables denoting the total weight of the minimum spanning tree, and the total weight of the power assignment corresponding to T_n, respectively.

The weight of the MST approximation of the power assignment problem depends on the number of *runs* of 1's. Here a run is defined as a succession of 1's preceded and succeeded by 0's. The number of elements in a run will be referred to as its *length*. Let R_{1i} denote the random variable indicating the number of runs of 1's of length i, and let R_{0i} denote the random variable indicating the number of runs of 0's of length i. In addition, let R_1 (resp R_0) be random variables indicating the total number of runs of 1's (resp 0s), N_1 (resp N_0) are random variables denoting the number of edges with weight 1 (resp weight 0). (Note that $R_1 \leq N_1$, $R_0 \leq N_0$, and $N_1 + N_0 = n$).

Example 2. (*Illustration of definition of runs.*) Let $n = 11$, so there are 10 edges. Both 0101010101 and 1111100000 are possible weight assignments with $W(T_n) = 5$. For the first series, the weight of the associated power assignment $W(P_n) = 10$ for the second series the associated power assignment has weight $W(P_{11}) = 6$. As defined above, for the number of runs of ones $R_{11} = 5$ and, for the number of runs of zero's, $R_{01} = 5$ for the first series and $R_{15} = 1$ and $R_{05} = 1$ for the second series.

We prove the following theorem:

Theorem 1. *Let $G = (V, E)$ be a graph on $n > 0$ vertices, where all elements of V are on a line, each edge $e \in E$ connects two neighboring vertices, and the cost $c(e)$ of edge $e \in E$ is 1 with probability p and 0 with probability $1 - p$. Then*

$$E[W(P_n)/W(T_n)] = 2 - p + \frac{1}{n}.$$

Proof. Clearly, $W(T_n) = N_1$. As each run of 1's of length i contributes with $(i+1)$ to the power assignment, we have:

$$W(P_n) = \sum_{i=1}^{n-1} R_{1i}(i+1) = \sum_{i=1}^{n-1} iR_{1i} + \sum_{i=1}^{n-1} R_{1i} = N_1 + R_1,$$

so in order to analyze $W(P_n)/W(T_n)$, it is sufficient to analyze

$$\frac{W(P_n)}{W(T_n)} = \frac{N_1 + R_1}{N_1} = 1 + \frac{R_1}{N_1} = 1 + U_1, \tag{6}$$

where U_1 is defined by: $U_1 = R_1/N_1$. The conditional distribution of R_1, given that $N_1 = n_1$, has been derived by Mood in [14].

$$P(R_1 = r_1 | N_1 = n_1) = \frac{\binom{n_1-1}{r_1-1}\binom{n-n_1+1}{r_1}}{\binom{n}{n_1}}. \tag{7}$$

For the expected number of runs, given $N_1 = n_1$, we find using (7):

$$E[R_1 | N_1 = n_1] = \sum_{k=1}^{n_1} k \frac{\binom{n_1-1}{k-1}\binom{n-n_1+1}{k}}{\binom{n}{n_1}} = \frac{(n - n_1 + 1)n_1}{n}. \tag{8}$$

We are interested in $E[U_1]$. Assume $n_1 > 0$. From (8) it follows that:

$$E[U_1|N_1 = n_1] = \frac{(n - n_1 + 1)}{n},$$

so that

$$\begin{aligned} E[U_1] &= \sum_{n_1=0}^{n} E[U_1|N_1 = n_1]P(N_1 = n_1) \\ &= \sum_{n_1=0}^{n} \frac{(n - n_1 + 1)}{n} p^{n_1}(1-p)^{n-n_1}\binom{n}{n_1} \\ &= \frac{1 + n - np}{n} = \frac{1}{n} + 1 - p, \end{aligned}$$

which by (6) completes the proof.

This result can be intuitively explained as follows. For large n, when p is very small, the runs are of length 1, in this case $W(P_n) = 2W(T_n)$. On the other extreme, when p is close to 1, most likely there is a single run of 1's, in which case $W(P_n) = W(T_n) + 1$.

3.2. Uniformly distributed weights

Next, we consider the situation where $G = (V, E)$ is a complete graph formed by $x_1, \ldots, x_n \in \mathbb{R}^1$ where $x_1 \leq \ldots \leq x_n$ and where in addition, the Euclidean distances between neighboring vertices x_i and x_{i+1} are independently and uniformly distributed in the interval $[0, 1]$. The distance-power gradient is denoted by α, so the transmit power threshold $c(e)$ of an edge $e = \{x_i, x_j\}$ is defined as follows:

$$c(x_i, x_j) = \text{dist}(x_i, x_j)^\alpha.$$

By removing vertices with zero distance to a neighbor, we may assume that all neighboring distances are strictly positive, therefore the MST is uniquely realized by the path $e_1, \ldots, e_{n-1}$, where each edge connects neighboring vertices x_i and x_{i+1} ($i = 1, \ldots, n-1$). As before, let $W(T_n)$, $W(P_n)$ denote the random variables corresponding to the total cost of the minimum spanning tree, and the total cost of the power assignment corresponding to T_n, respectively. The weight of the spanning tree is:

$$W(T_n) = \sum_{i=1}^{n-1} \text{dist}(x_i, x_{i+1})^\alpha.$$

In order to formulate our result, we introduce the notion of *convergence in probability* (see e.g. [11]). A sequence of random variables Y_n, which is dependent on n, converges in probability to the constant c (notation: $Y_n \xrightarrow{P} c$), if and only if, for every $\varepsilon > 0$, $\lim_{n\to\infty} P(|Y_n - c| < \varepsilon) = 1$. We call a sequence of random variables Y_n *with high probability* smaller than c (in notation $Y_n \leq_P c$) if for every $\varepsilon > 0, \lim_{n\to\infty} P(Y_n < c + \varepsilon) = 1$. We have the following lemma.

Lemma 2. *Let U_n and V_n be two random variables that converge in probability to the respective constants c and d. Then (a) the ratio U_n/V_n converges in probability to the constant c/d; (b) If, in addition, for each $n \in \mathbb{N}$ it holds that $U_n \leq V_n$ with probability 1. Then $U_n \leq_P d$.*

Proof. Omitted.

We say that the sequence X_n *converges in mean* towards X if $EX_n < \infty$ for all n and $\lim_{n\to\infty} E[|X_n - X|] = 0$ (see [11]). The lemma below combines two results relating convergence in probability to convergence in mean and vice versa.

Lemma 3. *Let U_n, V_n be random variables. Then (a) if U_n converges in mean to c, then $U_n \xrightarrow{P} c$. (b) If $V_n \xrightarrow{P} d$ and if $P(|V_n| \leq b) = 1$ for all n and some $b \in \mathbb{R}$, then V_n converges in mean to d.*

Proof. Omitted.

We will use these facts as follows.

Lemma 4. *Let U_n and V_n be sequences of random variables, where U_n converges in mean towards $c \in \mathbb{R}$ and V_n converges in mean towards $d \in \mathbb{R}$. Moreover, assume there exist $p, q \in \mathbb{R}$ so that for all $n \in \mathbb{N}$ it holds that, $p \leq U_n/V_n \leq q$. Then the random variable U_n/V_n converges both in mean and in probability to c/d.*

Proof. By Lemma 3(a) we have $U_n \xrightarrow{P} c$ and $V_n \xrightarrow{P} d$. By Lemma 2(a) also $U_n/V_n \xrightarrow{P} c/d$ (this shows the last assertion of the lemma). As we assumed U_n/V_n to be bounded from above and below we can apply Lemma 3 (b) to conclude that U_n/V_n converges in mean to c/d.

Let $X_{(1)}, X_{(1)}, \ldots, X_{(1)}$ denote the order statistics of the random sample $X_1, \ldots, X_n$.

We use the following formula for order statistics on n variables, derived in [6], for $\alpha \in \mathbb{N}$.

$$E[X_{(r)}^{\alpha}] = \frac{n!}{(n+\alpha)!}\frac{(r-1+\alpha)!}{(r-1)!}. \tag{9}$$

For $k \in \mathbb{N}$, we have the following identity which is easily proved by induction:

$$\sum_{r=1}^{k} \frac{(r+a)!}{r!} = \frac{(1+a+k)!}{(1+a)k!} - a! \tag{10}$$

Now we are in a position to formulate the main result of this section.

Theorem 2. *Let $G = (V, E)$, $|V| = n$, be the complete graph formed by $x_1, \ldots, x_n \in \mathbb{R}^1$, where the distance between neighboring vertices is independently uniformly distributed on the interval $[0,1]$ and the power cost $c(e)$ of an edge $e = \{x_i, x_j\}$ depends on the distance as: $c(e) = dist(x_i, x_j)^{\alpha}$, where $\alpha \in \mathbb{N}$. Then*

$$\lim_{n\to\infty} \frac{W(P_n)}{W(T_n)} \leq_P 2 - 2^{-\alpha},$$

and

$$\lim_{n\to\infty} E[\frac{W(P_n)}{W(T_n)}] \leq 2 - 2^{-\alpha}$$

Proof. To simplify notation, we assume n is even. (The proof for n is odd is identical to the proof presented below, except that in many occasions $n/2$ needs to be replaced by either $\lfloor n/2 \rfloor$, or $\lceil n/2 \rceil$). We may assume that the minimum spanning tree is realised by $e_1, \ldots, e_{n-1}$ where each edge connects neighboring vertices $e_i = \{x_i, x_{i+1}\}$ $(i = 1, \ldots, n-1)$. Let $X_1, \ldots, X_{n-1}$ denote the random variables corresponding to the (uniformly distributed) distances $\text{dist}(i, i+1)$, $(i = 1, \ldots, n-1)$ and let $X_{(i)}$ denote the i-th order statistic of the random sample $X_1, \ldots, X_{n-1}$. (Note that raising the variables X_i to a positive power α maintains the order of the variables.)

Next, define the average of the highest $n/2$-values of the random sample $X_1^{\alpha}, \ldots, X_{n-1}^{\alpha}$:

$$Y_n = \frac{2}{n}\sum_{i=n/2}^{n-1} X_{(i)}^{\alpha}.$$

Similarly, we define the average value of the complete random sample.

$$Z_n = \frac{1}{n}\sum_{i=1}^{n-1} X_{(i)}^{\alpha} = \frac{1}{n}\sum_{i=1}^{n-1} X_i^{\alpha}.$$

Clearly $W(T_n) = nW(Z_n)$ and by inequality (3) we have that: $W(P_n) \leq nY_n$. So

$$\frac{W(P_n)}{W(T_n)} \leq \frac{W(Y_n)}{W(Z_n)}. \tag{11}$$

Moreover, observe that $1 \leq Y_n/Z_n \leq 2$.

By (11), Lemma 2 and Lemma 4, it is sufficient to show that:

$$\mu_Y = \lim_{n\to\infty} EY_n = \frac{2 - 2^{-\alpha}}{\alpha + 1} \tag{12}$$

and

$$\mu_Z = \lim_{n\to\infty} E[Z_n] = \frac{1}{\alpha + 1}. \tag{13}$$

Clearly, the division μ_Y/μ_Z yields the desired ratio. To see (12), it follows from (9) and (10) that

$$E[Y_n] = \frac{2}{n} E\left[\sum_{r=n/2}^{n-1} X_{(r)}^{\alpha}\right] = \frac{2}{n} \frac{n!}{(n+\alpha)!} \sum_{r=n/2}^{n-1} \frac{(r-1+\alpha)!}{(r-1)!} = \tag{14}$$

$$\frac{2}{\alpha+1}\left(\frac{n-1}{\alpha+n} - \frac{(\alpha+n/2-1)!(n-1)!}{(n/2-2)!(n+\alpha)!}\right).$$

To simplify the expression for $E[Y_n]$, we define $u_n(\alpha)$ as:

$$u_n(\alpha) = \frac{(\alpha+n/2-1)!(n-1)!}{(n/2-2)!(n+\alpha)!}.$$

We show by induction on α:

$$\lim_{n\to\infty} u_n(\alpha) = \frac{1}{2^{\alpha+1}}. \tag{15}$$

For the base case $\alpha = 0$, we have:

$$u_n(0) = \frac{(n/2-1)!(n-1)!}{(n/2-2)!n!} = \frac{n/2}{n} = \frac{1}{2},$$

as required. Now suppose (15) has been proven for integers $1, \ldots, \alpha$. Then for $\alpha + 1$ we obtain:

$$u_n(\alpha+1) = \frac{(\alpha+1+n/2-1)!(n-1)!}{(n/2-2)!(n+\alpha+1)!} = \frac{\alpha+n/2}{n+\alpha+1} u_n(\alpha)$$

So for $n \to \infty$ we have:

$$\lim_{n\to\infty} u_n(\alpha+1) = \lim_{n\to\infty} \frac{1}{2} u_n(\alpha) = \frac{1}{2^{\alpha+2}},$$

showing (15).

Substituting the expression for $u_n(\alpha)$ in (14) we obtain:

$$\mu_Y = \lim_{n\to\infty} E[Y_n] = \frac{2-2^{-\alpha}}{\alpha+1},$$

as desired.

To see (13), it follows from (9) and the fact that $EX_i^\alpha = 1/(\alpha+1)$ that:

$$E[Z_n] = \frac{1}{n}E\left[\sum_{i=1}^{n-1} X_{(i)}^\alpha\right] = \frac{1}{n}E\left[\sum_{i=1}^{n-1} X_i^\alpha\right] = \frac{1}{n}\sum_{i=1}^{n-1} E[X_i^\alpha] = \frac{n-1}{n(\alpha+1)},$$

whence

$$\mu_Z = \lim_{n\to\infty} E[Z_n] = \frac{1}{\alpha+1}.$$

This finishes the proof.

4. General case

In this section we generalize our analysis to higher dimensions, where we make use of the results from [17] to bound the ratio $W(P_n)/W(T_n)$. Consider n points (denoted $\eta_1,\ldots,\eta_n$) random, independently uniformly distributed on the d-dimensional unit cube $B = \left(-\frac{1}{2},\frac{1}{2}\right]^d$. H_n is the point process $\eta_1,\ldots,\eta_n$. Note here that an essential difference with Section 3, is that here a large number of vertices is distributed in a bounded region, whereas in Section 3 we analyzed the behavior of the algorithm in case the region was not bounded.

Often, in conjunction with H_n also P_n is considered. Here P_n denotes the Poisson point process $P_n = \{\eta_1,\ldots,\eta_{N_n}\}$ where N_n is a Poisson variable with mean n independent of $\{\eta_i\}$. So P_n is simply a homogenous Poisson process on the cube of rate n.

To eliminate boundary effects as discussed in [21], the toroidal model is considered. In this model, instead of the Euclidean metric $(\text{dist}(\eta_i,\eta_j))$, we use the metric

$$\text{tdist}(\eta_i,\eta_j) = \min_{z\in\mathbb{Z}^d} \|\eta_i-\eta_j-z\|.$$

The Nearest Neigborhood Graph (NNG) is the graph where each point is connected to its nearest neighbor. Note that NNG $\subset$ MST. As in [17], we call an edge e of the MST or NNG σ-long if

$$n\pi_d\|e\|^d - \log n > \sigma \tag{16}$$

Here, π_d denotes the volume of the unit ball in d dimensions ($\pi_d = \pi^{d/2}/\Gamma((d/2)+1)$).

As this is the basis of our work, we also formulate the main result of [17].

Theorem 3. (Penrose [17].) *Consider the toroidal model with $d \geq 2$ or the Euclidean model with $d = 2$. Let $\sigma \in \mathbb{R}$. Then with probability approaching 1 as $n \to \infty$, every σ-long edge of the MST on P_n or on H_n is also in the corresponding NNG, and moreover, every such edge has an end at a leaf (vertex of degree 1) of the MST.*

This theorem implies that both the MST and NNG contain the same number of σ-long edges. (Each σ-long edge of the MST is, according to Theorem 3, also in the NNG, and as NNG $\subset$ MST the converse is also true.)

According to [17] (cf. Lemma 2 and below, page 345), the following holds for the number of edges of the NNG and MST:

Lemma 5. *For the toroidal model with $d \geq 1$ or the Euclidean model with $d \leq 2$, the asymptotic distribution as $n \to \infty$ of the number of σ-long edges of the Nearest Neighborhood Graph is Poisson with mean $e^{-\sigma}$.*

For an MST edge e, let us denote by $\lambda(e)$ the *rescaled* length of e according to (16). That is,

$$\lambda(e) = n\pi_d \|e\|^d - \log n. \tag{17}$$

Let MST be a minimum spanning tree on H_n. By the random variable n_σ we denote the number of MST edges e that is σ-long, i.e., for which $\lambda(e) > \sigma$. By Lemma 5 we have

$$P(n_\sigma = k) = \frac{e^{-\sigma k}}{k!} e^{-e^{-\sigma}} \quad k = 0, 1, 2, \ldots \tag{18}$$

For future reference we will present here some useful properties of the Γ-function, where we consider only the Γ-function with real arguments ($x \in \mathbb{R}$). The Γ-function is defined as:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} \mathrm{d}t. \tag{19}$$

For the derivative it holds that:

$$\Gamma'(x) = \int_0^\infty t^{x-1} e^{-t} \log(t) \mathrm{d}t. \tag{20}$$

The digamma function is defined as

$$\psi(x) = \Gamma'(x)/\Gamma(x). \tag{21}$$

For integer arguments, it is well-known (see [1], equation 6.3.2) that

$$\psi(k) = -\gamma + H(k-1) \;\; k = 1, 2, \ldots, \tag{22}$$

where γ denotes the Euler-constant ($\gamma = \lim_{m\to\infty} H(m) - \log m \approx 0.577216$) and $H(m)$ denotes the $m-$th harmonic number. (Where, by definition, $H(0) = 0$, and for $m \geq 1$: $H(m) = \frac{1}{1} + \ldots + \frac{1}{m}$.) From this, we obtain the following identity for the sum of $\psi(.)$ with integer arguments.

Lemma 6.

$$\sum_{k=1}^{s} \psi(k) = s(\psi(s) - 1) + 1 \tag{23}$$

Proof. We find by applying (22) for the first and third equality:

$$\sum_{k=1}^{s} \psi(k) = -s\gamma + \sum_{k=1}^{s} H(m-1) = -s\gamma + sH(s-1) - (s-1) = s\psi(s) - (s-1),$$

where the second equality follows from the general identity (easily proved by induction):

$$\sum_{k=1}^{s} H(k-1) = sH(s-1) - (s-1).$$

This finishes the proof.

In the limit we have, (see [1], equation 6.3.18):

$$\lim_{x\to\infty} \psi(x) - \log(x) = 0. \tag{24}$$

In Section 4.3, we will use the incomplete Γ-function which is defined as:

$$\Gamma(a, x) = \int_x^\infty t^{a-1} e^{-t} \mathrm{d}t.$$

Next, we consider the edge lengths of a MST on H_n as random variables $X_0, \ldots, X_{n-2}$. To keep notation simple, we number the order statistics $X_{(0)}, \ldots, X_{(n-2)}$, so that $X_{(0)} \geq X_{(1)} \geq \ldots \geq X_{(n-2)}$. With this notation: $\lambda(X_{(m)}) \leq \sigma$ if and only if $n_\sigma \leq m$, $m = 0, \ldots, n-2$.

Lemma 7. *Let T_n be a minimum spanning tree on H_n. The probability distribution of the rescaled length $\lambda(X_{(m)})$ as defined in (11) of the m-th longest edge $X_{(m)}$ of T_n is for $m = 0, \ldots, n-2$ defined by:*

$$P(\lambda(X_{(m)}) \leq \sigma) = \sum_{k=0}^{m} \frac{\exp(-\sigma)^k}{k!} e^{-\exp(-\sigma)} = \frac{\Gamma[m+1, e^{-\sigma}]}{m!}. \tag{25}$$

With probability density function f_m given by:

$$f_m(\sigma) = \frac{(e^{-\sigma})^{(m+1)} e^{-(e^{-\sigma})}}{m!}, \tag{26}$$

and expected value

$$E[\lambda(X_{(m)})] = \int_{-\infty}^{\infty} \sigma f_m(\sigma) d\sigma = -\frac{1}{m!} \int_0^{\infty} t^m e^{-t} \log(t) dt = -\psi(m+1). \tag{27}$$

Proof. The m-th longest edge has rescaled length $\leq \sigma$ if and only if it is not σ-long. This means that the number of σ-long edges is less than (or equal to) m. Now observe that $\lambda(X_{(m)})$ has a gamma distribution with parameters $(e^{-\sigma}, m)$, and thus density (26). The expected value follows by integration over the real numbers, where we use (20) and (21). The second equality of (27) follows by substitution $t = e^{-\sigma}$.

4.1. Two dimensions and distance power gradient α =2

We assume $d = 2$ and the distance power gradient $\alpha = 2$. We prove the following theorem:

Theorem 4. *Let $G = (V, E)$ be a complete graph formed by n points (denoted $\eta_1, \ldots, \eta_n$) in $\mathbb{R}^2$ where the η_i are random, independently uniformly distributed on the unit cube $B = \left(-\frac{1}{2}, \frac{1}{2}\right]^2$, and the cost $c(e)$ of an edge $e = \{\eta_i, \eta_j\}$ is dist$(\eta_i, \eta_j)^2$. Then,*

$$\lim_{n\to\infty} \frac{W(P_n)}{W(T_n)} \leq_P 1 + \log 2, \tag{28}$$

and

$$\lim_{n\to\infty} E[\frac{W(P_n)}{W(T_n)}] \leq 1 + \log 2. \tag{29}$$

Proof. To simplify notation, we assume n is even. (The proof for odd n goes along the same lines, with $n/2$ replaced by $\lfloor n/2 \rfloor$, or $\lceil n/2 \rceil$.) Let $X_0, \ldots, X_{n-2}$ denote the random variables corresponding to lengths of T_n. Let $X_{(i)}$ denote the i-th order statistic of the random sample $X_0, \ldots, X_{n-2}$. Again, the numbering is chosen so that $X_{(0)} \geq X_{(1)} \geq \ldots \geq X_{(n-2)}$. Next, define the sum of the highest $n/2$-values of the random sample $X_0, \ldots, X_{n-2}$.

$$Y_n = \sum_{i=0}^{n/2-1} X_{(i)}^2$$

Similarly, we define the sum of the complete random sample.

$$Z_n = \sum_{i=0}^{n-2} X_{(i)}^2$$

Again, we define $\mu_Y = \lim_{n\to\infty} E[Y_n]$ and $\mu_Z = \lim_{n\to\infty} E[Z_n]$. Clearly $W(T_n) = W(Z_n)$ and by by inequality (5) we have that: $W(P_n) \leq 2Y_n$. Further, we observe that $1/2 \leq Y_n/Z_n \leq 1$ for all $n \in \mathbb{N}$.

By Lemma 2 and Lemma 4, it is sufficient to show that:

$$\mu_Y = \lim_{n\to\infty} E[Y_n] = \frac{1+\log 2}{2\pi} \tag{30}$$

and

$$\mu_Z = \lim_{n\to\infty} EZ_n = \frac{1}{\pi}. \tag{31}$$

Clearly, the division μ_Y/μ_Z yields the desired ratio (taking into account that $W(P_n) \leq 2Y_n$). To see (30), we note that for $\alpha = 2$, (17) reads:

$$\lambda(e) = n\pi\|e\|^2 - \log n \tag{32}$$

By equation (27) we obtain for the expected value of $\lambda(X_{(m)}$:

$$E[\lambda(X_{(m)})] = -\psi(m+1) \ \ (m = 0, \ldots, n-2).$$

By (32) and the fact that n, π are constants we find for the square of the length of the m-th longest edge $(m = 0, \ldots, n-2)$:

$$E[X^2_{(m)}] = \frac{-\psi(m+1) + \log n}{n\pi} \tag{33}$$

For the expected value of the sum of the $n/2$ longest edges we find therefore, using (23) :

$$\begin{aligned}
\sum_{m=0}^{n/2-1} E[X^2_{(m)}] &= \sum_{m=0}^{n/2-1} \frac{-\psi(m+1) + \log n}{n\pi} \\
&= \frac{1}{n\pi} \sum_{m=1}^{n/2} (\log n - \psi(m)) \\
&= \frac{(n/2)}{n\pi}(\log n - \psi(n/2) + 1)) + \frac{1}{n\pi}
\end{aligned}$$

So we obtain by using (24) and the fact that $\log x = \log(x/2) + \log 2$,

$$\begin{aligned}
\mu_Y &= \lim_{n\to\infty} E[Y_n] \\
&= \frac{1}{2\pi} \lim_{n\to\infty} [\log n - \psi(n/2) + 1] + \lim_{n\to\infty} \frac{1}{n\pi} \\
&= \frac{1}{2\pi} \lim_{n\to\infty} [\log 2 + \log(n/2) - \psi(n/2) + 1] \\
&= \frac{1 + \log 2}{2\pi},
\end{aligned}$$

as required.

To see (31) we note that for the expected value $E[Z_n]$ of the sum of the square lengths of the minimum spanning tree edges it follows from (32) and (23) that:

$$\begin{aligned}\sum_{m=0}^{n-2} E[X_{(m)}^2] &= \sum_{m=0}^{n-2} \frac{-\psi(m+1)+\log n}{n\pi} \\ &= \frac{1}{n\pi}\left((n-1)\log n - \sum_{m=1}^{n-1}\psi(m)\right) \\ &= \frac{(n-1)}{n\pi}\left(\log n - \psi(n-1)+1\right) + \frac{1}{n\pi}\end{aligned}$$

So we obtain using (24) ,

$$\mu_Z = \lim_{n\to\infty} E[Y_n] = \frac{1}{\pi}\lim_{n\to\infty}\left[\frac{n-1}{n}\left(\log n - \psi(n-1)+1\right) + \frac{1}{n}\right] = \frac{1}{\pi}, \qquad (34)$$

as required.

This finishes the proof.

4.2. d dimensions and distance power gradient $\alpha = d$.

In fact, the analysis we presented above for the quadratic power model in 2 dimensions goes through with minor changes to derive the same result for a distance power gradient $\alpha = d$ in d dimensions. More specifically,

Theorem 5. *Let $G = (V, E)$, $|V| = n$, be a complete graph formed by $\eta_1, \ldots, \eta_n \in \mathbb{R}^d$ independently uniformly distributed on the unit cube $B = (-\frac{1}{2}, \frac{1}{2}]^d$, where the cost $c(e)$ of an edge $e = \{x_i, x_j\}$ is tdist$(x_i, x_j)^d$, where 'tdist' denotes the distance according to the Toroidal model. Then,*

$$\lim_{n\to\infty} \frac{W(P_n)}{W(T_n)} \leq_P 1 + \log 2, \qquad (35)$$

and

$$\lim_{n\to\infty} E[\frac{W(P_n)}{W(T_n)}] \leq 1 + \log 2. \qquad (36)$$

Proof. Using (27) we find by using the fact that n, π are constants :

$$E[X_{(m)}^d] = \frac{-\psi(m+1)+\log n}{n\pi_d}. \qquad (37)$$

The rest of the proof is the same as for Theorem 4.

4.3. General case: $\alpha \neq d$

For completeness, we also present the formula's for the cases where $\alpha \neq d$.

Lemma 8. *Let T_n be a minimum spanning tree on H_n, where the 'length' is measured according to the Euclidean distance if $d = 2$ and according to the toroidal distance if $d > 2$. Then the probability distribution of the length $X_{(m)}$ of the m-th longest edge $X_{(m)}$ of T_n is for $m = 0, \ldots, n-2$ defined by:*

$$P(X_{(m)} \leq \beta) = \frac{\Gamma(1+m, e^{-\beta^d n\pi_d + \log n})}{\Gamma(1+m)} = \frac{\Gamma(1+m, ne^{-\beta^d n\pi_d})}{\Gamma(1+m)} \tag{38}$$

where $\beta \geq 0$. For the associated probability density function g_m we obtain

$$g_m(\beta) = \pi_d \beta^{d-1} dn^2 e^{-n(e^{-\beta^d n\pi_d} + \beta^d \pi_d)} \left(ne^{-\beta^d n\pi_d}\right)^m. \tag{39}$$

For the expected length of $X_{(m)}$ it holds that

$$E\left[g_m(\beta)\right] = \int_0^\infty \pi_d \beta^d dn^2 e^{-n(e^{-\beta^d n\pi_d} + \beta^d \pi_d)} \left(ne^{-\beta^d n\pi_d}\right)^m d\beta. \tag{40}$$

Proof. Equation (38) follows directly from the probability distribution of the rescaled lengths. The probability density function g_m is obtained by differentiating with respect to β, and the expectation is found by integrating βg_m over $\mathbb{R}^+$.

With these equations in principle the performance of the minimum spanning tree heuristic for the power assignment problem could be analysed for the cases where $\alpha \neq d$. However, this situation is more complex than the case where $\alpha = d$. For example, it follows from (38) that for fixed $\beta > 0$ and $m \in \mathbb{N}$:

$$\lim_{n\to\infty} P(X_{(m)} \leq \beta) = 1 \tag{41}$$

5. Conclusions and further research

We have presented an average case analysis of the ratio P_n/T_n which provides an upper bound for the ratio P_n to the value optimal power assignment. The strategy used, is to first bound the P_n in terms of the 'longest' edges of T_n, and then use formula's for the longest minimum spanning tree edges. Extension of this strategy to the minimum spanning tree where all edges are uniformly distributed is straightforward by a theorem of Frieze [7]. Concerning the MST heuristic, it would be interesting to further investigate the case where $\alpha \neq d$. Even more interesting would be to investigate heuristics as presented in [2].

Acknowledgements

This work was supported under the Casimir grant of National Science Institute (N.W.O.). Thanks to Tom Coenen who pointed out [17] to us.

References

[1] Abramowitz, M. and Stegun, I.A. (1972). *Handbook of mathematical functions"*, National Bureau of Standards, Applied Mathematics Series–55, Tenth Printing. U.S. Government Printing Office, Washington D.C. 20402.

[2] Althaus, E., Calinescu, G., Mandoiu, I.I. , Prasad, S., Tchervenski, N., Zelikovsky, A. (2006). Power efficient range assignment for symmetric connectivity in static ad hoc wireless networks, *Wireless Networks*, **12,** 287–299.

[3] Blough, D. (2002). On the Symmetric Range assignment problem in wireless ad-hoc networks, In *Proceedings of the 2nd IFIP International Conference on Theoretical Computer Science (TCS).*

[4] Chen, W. and Huan, N. (1989). The strongly connecting problem on multihop packet radio networks, *IEEE Transactions on Communications*, **37,** 293- 295.

[5] Clementi, A.E.F, Penna, P. and Silvestri, R. (2000). On the power assignment problem in radio networks, *Electronic Colloquium on Computational Complexity*, Report No. 54.

[6] David, F.N. and Johnson, N.L. (1954), Statistical treatment of censored data, I. Fundamental Formulae,*Biometrika*, **41**, 228–240.

[7] Frieze, A.M. (1985). On the value of a random minimum spanning tree problem. *Discrete Applied Mathematics*, **10,** 47–56.

[8] FEENEY, L.M. AND NILSON, M. (2001). Investigating the energy consumption of a wireless network interface in an ad-hoc networking environment, *Proc. 20th IEEE INFOCOM*, 1548–1557.

[9] FUCHS, B. (2006). On the hardness of range assignment problems, *Algorithms and Complexity*, Lecture Notes in Computer Science, Springer Berlin, Heidelberg, **3998,** 127–138.

[10] GRIMETT, G.R. AND STIRZAKER, D.R., Probability and Random Processes, 2nd Edition, Clarendon Press, Oxford, pp 271–285, ISBN 0-19-853665-8.

[11] HOGG, R.V. CRAIG, A.T, Introduction to mathematical statistics, 4th edition.

[12] KIROUSIS L., KRANAKIS E. , KRZANC, D., PELC, A. (2000). Power consumption in packet radio networks, *Theoretical Computer Science* **243,** 289–205

[13] LLOYD, E., LIU, R., MARATHE, M., RAMANATHAN, R., RAVI, S. (2005). Algorithmic Aspects of Topology Control problems for ad-hoc networks, *Mobile Networks and applications*, **10**, Issue 1-2 , 19–34.

[14] MOOD, A. M. (1940). The Distribution Theory of Runs, *The Annals of Mathematical Statistics* **11,** No. 4, 367–392.

[15] MONTEMANNI, R. AND GAMBARDELLA, L.M. (2005). Exact algorithms for the minimum Power symmetric connectivity problem in wireless networks, *Computers and Operations Research*, **32**, Issue 11, 2891–2904.

[16] PAHLAVAN, K. AND LEVESQUE, A. (1995). *Wireless Information Networks*, Wiley-Interscience, 1995.

[17] PENROSE, M. D. (1997). The Longest Edge of the Random Minimal Spanning Tree, *The Annals of Applied Probability* , **7,** 340–361

[18] RAMANATHAN, R. AND ROSALES-HAIN, R. (2000). Topology control of multihop wireless networks using transmit power adjustment, in: *Proc. IEEE INFOCOM*, 404–413.

[19] RODOPLU, V. AND MENG, T.H. (1999) Minimum energy mobile wireless networks, *IEEE J. Select. Areas Communications*, **17,** no. 8, 1333–1344.

[20] Santi,P. Blough, D., Vainstein,F. (2001). A probabilistic analysis for the range assignment Problem in ad-hoc networks, in: *MobiHoc '01: Proceedings of the 2nd ACM International symposium on Mobile ad hoc networking & computing*, ACM Press, New York, 212–220.

[21] Steele, J. M. and Tierney, L. (1986). Boundary domination and the distribution of the largest nearest-neighbor link in higher dimensions, *Journal of Applied Probability*, **23,** 524–528.

[22] Steele, J.M. (1988). Growth rates of Euclidean minimal spanning trees with power weighted edges, *Ann. Probabil.*, **16,** 1767–1787.

[23] Wattenhofer, R., Li, L., Bahl, V. and Wang, Y.M. (2001). Distributed topology control for power efficient operation in multihop wireless ad-hoc networks, in: *proc. twentieth Annual joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, 1388–1397.