The Non-uniform Bounded Degree Minimum Diameter Spanning Tree Problem with an Application in P2P Networking

Jakarin Chawachat[a,*], Jittat Fakcharoenphol[a], Wattana Jindaluang[a]

[a]*Department of Computer Engineering, Kasetsart University, Bangkok 10900, Thailand*

Abstract

This paper considers the Bounded Degree Minimum Diameter Spanning Tree problem (BDST problem) with non-uniform degree bounds. In this problem, we are given a metric length function ℓ over a set V of n nodes and a degree bound B_v for each $v \in V$, and want to find a spanning tree with minimum diameter such that each node v has degree at most B_v. We present a simple extension of an $O(\sqrt{\log n})$-approximation algorithm for this problem with uniform degree bounds of Könemann, Levin, and Sinha (Könemann *et al.*, 2003) to work with non-uniform degree bounds. We also show that this problem has an application in the peer-to-peer content distribution. More specifically, the Minimum Delay Mesh problem (MDM problem) introduced by Ren, Li and Chan (Ren *et al.*, 2008) under a natural assumption can be reduced to this non-uniform case of the BDST problem.

Key words: Approximation algorithms, Graph theory, Degree bounds, Spanning trees.

1. Introduction

We study the *Bounded Degree Minimum Diameter Spanning Tree* problem (BDST problem) with non-uniform degree bounds. Given a metric length function ℓ over a set V of n nodes, and a degree bound B_v for each $v \in V$, the BDST problem with non-uniform degree bounds is to find a spanning

*Corresponding author

Email addresses: `cjakarin@gmail.com` (Jakarin Chawachat), `jittat@gmail.com` (Jittat Fakcharoenphol), `w.jinda@chiangmai.ac.th` (Wattana Jindaluang)

tree T of G of minimum diameter such that each node v has degree at most B_v in T.

In this paper, we show that the Könemann, Levin, and Sinha's algorithm [1] which works for the case that degree bound $B_v = B$ for all nodes v, can be easily extended to handle the non-uniform case. One of our main tools is the result of Frederickson, Hecht, and Kim [2] for the Minimum Metric Traveling k-Salesperson Problem.

The following theorem is the main result of this paper.

Theorem 1. *There is a polynomial-time $O(\sqrt{\log n})$-approximation algorithm for the BDST with non-uniform degree bounds on n nodes.*

An application in peer-to-peer (P2P) networking which is our motivation is discussed in Section 2. Related work is presented in Section 3. Section 4 describes the main result.

2. Application to P2P networking

As an application to P2P networking, we consider the *Minimum Delay Mesh problem* (MDM problem), introduced by Ren, Li, and Chan [3] who proved the NP-Hardness and gave a heuristic algorithm. In this problem, we are given an overlay network as a complete directed graph $G = (V, E)$, where V is a set of peers and E is a set of overlay links. Each edge (u, v) has a cost $\ell(u, v)$ corresponding to the delay from node u to node v in the network. The delay of path P is the sum of the delay on each edge in P, i.e., $\sum_{(u,v)\in P} \ell(u, v)$.

In P2P networks, each node can also forward the information to other nodes. For each node u, we denote an uplink bandwidth of u as U_u.

Let *unit* be the minimum package size in the streaming application. We assume that there is only one source $r \in V$. Let s be the streaming rate, i.e., r distributes the content at the rate of s units per second. The goal of this problem is to find a streaming scheme so that each peer receives at least s units per second.

A *streaming scheme* describes how the stream is distributed to each peer from the source. It assigns, for each peer u, a set of parents P_u that would relay parts of the streams to u. The amount of stream bandwidth u receives from v is denoted by $s_u(v)$.

The scheme must respect the uplink bandwidth, i.e., $\sum_{u:P_u \ni v} s_u(v) \leq U_v$ for each parent v. Also, the parent-child relationship defined by P_u's for every $u \in V$ must be acyclic.

Given the problem instance, the MDM problem is to find a streaming scheme that respects the uplink requirement of every node while minimizing the maximum delay from source r to any node.

This paper considers a special case of the MDM problem where all bandwidths are multiples of the required bandwidth s. In this case, we show that there exists an optimal tree solution in this special case. We assume that the delays form a metric, i.e., the delays are symmetric and satisfy the triangle inequality. Without loss of generality, we adjust all bandwidths by dividing by s and then round off the values.

Lemma 1. *If all bandwidths are multiples of the required bandwidth s and the delays are metric, any MDM solution can be converted to a solution that forms a tree whose diameter is no greater than the original solution.*

Proof. We shall use the max-flow min-cut theorem to construct the tree. To make the source-sink of the flow in our proof explicit, we shall construct a bipartite graph that represents the streaming schemes.

Given the solution of the MDM problem, let a bipartite graph $F = (V_1 \cup V_2, E_F)$ be such that $V_1 = V_2 = V$ and for each pair of nodes u and v such that $u \in P_v$ and $s_v(u) > 0$, we have $(u, v) \in E_F$ where $u \in V_1$ and $v \in V_2$, i.e., we have an edge (u, v) in F iff u provides v with the stream.

To construct a tree solution, it suffices to show that we can find, for each $u \neq r$, a parent P_u satisfying the following two conditions: (1) for each u, $(P_u, u) \in E_F$, and (2) each node v has at most U_v children. Since this parent assignment for each node induces a spanning tree, the lemma follows.

We present an algorithm to find the assignment. We reduce the problem to the max-flow problem using a standard reduction by adding a super source a and a super sink b. We give each edge (u, v) in F a capacity of the smallest integer not less than the amount of direct download of v from u in a given solution. We then direct all edges in F from V_1 to V_2. We add a directed edge (a, u) with capacity U_u for each u in V_1. We also add a directed edge (v, b) with capacity 1 for each v in V_2. One can verify that the existence of the given solution to the MDM problem implies that the size of the minimum ab cut is at least $|V| - 1$. Therefore, from the Max-Flow Min-Cut theorem, we know that a flow of value $|V| - 1$ exists. Since all capacities are integers, an integral flow also exists. Given the integral flow, each node in $V_2 - \{r\}$

receives a unit flow from node in V_1 which becomes its parent. Then, one can extract the required assignment. □

With this fact, the problem reduces to the non-uniform case of BDST problem by assigning the degree bound $B_v = U_v + 1$ for each node $v \in V - \{r\}$, and assigning the degree bound $B_r = U_r$ for the source r. Then, we have the following corollary.

Corollary 1. *If all bandwidths are multiples of the required bandwidth s, there is an $O(\sqrt{\log n})$ approximation algorithm for the MDM problem.*

3. Related work

In [1], Könemann, Levin, and Sinha proposed an approximation algorithm for the BDST problem with uniform degree bounds. The algorithm of Könemann, Levin, and Sinha (KLS algorithm) first partitions the graph into low-diameter clusters. For each cluster, they build a balanced $(B-1)$-ary tree. After that, they construct a backbone tree over the clusters. Then, their algorithm gave an $O(\sqrt{\log_B n})$-approximation algorithm when $B \geq 3$. They also proved the BDST problem is NP-hard.

Ravi [4] studied a closely-related problem called the Minimum Broadcast Time problem, which given a connected undirected graph and a rooted node, one wants to find a scheme that accomplishes the broadcast in the minimum number of steps. In each step, each node is allowed to send the message to at most one of its neighbors. The major difference between this problem and our problem is that this problem considers the non-metric case. They use the notion of poise to find the minimum broadcasting time. The poise of a spanning tree is the sum of the diameter of the tree and the maximum degree of any node. The poise of a graph is the minimum poise of its spanning trees. They give an $O(\frac{\log^2 n}{\log\log n})$-approximation algorithm for the Minimum Broadcast Time problem on n node. Later, Elkin and Kortsarz [5] improved the bound of this problem to $O(\frac{\log n}{\log\log n})$.

Recently, Huang, Ravindran, and Kumar [6] studied the MDM problem (under a different name, the Minimum Delay Peer-to-Peer Streaming problem) and used similar approach to ours, i.e., they reduced the problem to some form of the BDST problem and used a modified version of the KLS algorithm.

They considered another set of assumptions to the MDM problem, i.e., (1) at least half of the nodes have uplink bandwidth at least $2s$ units/second,

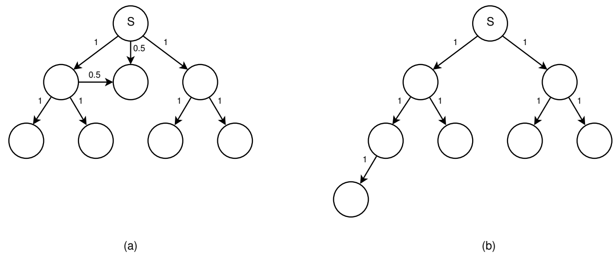

Figure 1: shows that the optimal solution is not always a tree. (a) mesh and (b) tree.

and (2) all nodes must have uplink bandwidth more than 1 unit/second. They claimed that under these assumptions, the optimal streaming scheme is a tree or a combination of multiple trees ([6] section 4.2.4).

We shall show that the assumptions above are not sufficient to ensure that the optimal solution is a tree. Consider a network with 8 nodes where each node has an uplink bandwidth 2.5 units/second, the required bandwidth is 1 unit/second, and each edge has a unit delay. With this setting, we found that the minimum delay in any tree is 3 units but the minimum delay in the given mesh is 2 units (see Figure 1). This example contradicts their claim.

It is interesting to see if the one can still use the tree solutions to approximate this general case.

4. An $O(\sqrt{\log n})$ approximation algorithm for the non-uniform BDST problem

In order to prove our main result, we will assume at first that there exists a node v with a degree bound $B_v > 2$, which we will choose as a root node r. We shall remove this assumption in Section 4.5.

Let Δ denote the optimal diameter. As in [1], we make an assumption that the algorithm knows Δ because it can find the appropriate value by binary search.

Our algorithm finds a spanning tree T rooted at r where the distance from r to any node in T is at most $O(\sqrt{\log n})\Delta$. To construct the solution, we partition V into 3 sets V_1, V_2, and V_3 as follows.

- For $i \in \{1, 2\}$, V_i is the set of nodes v with $B_v = i$.
- V_3 is the set of nodes v with $B_v > 2$. (Note that, $r \in V_3$.)

For nodes in V_3, we construct a tree H using the Könemann *et al.*'s algorithm [1]. We later call this tree the *base* of the solution. For nodes in V_2, we construct a set of paths, $L_1, L_2, \ldots$, using the Frederickson *et al.*'s algorithm [2]. These paths are called the *tails* of the solution. Then we attach each path L_i to nodes in H. Finally, we attach each node in V_1, which can only be a leaf node, to the end of each tail.

Before we get into the core material, we provide reviews on properties of trees and discuss briefly the result by Frederickson, Hecht, and Kim [2].

4.1. Preliminary

4.1.1 Properties of trees. For any tree $T = (V, E)$ and node $u \in V$, we denote by $deg_T(u)$ the degree of u. The nodes in T whose degrees are 1 are called *leaves*. All non-leaf nodes are called *internal nodes*. We denote by $N(T)$ the set of internal nodes of T. Most of the following propositions follows from standard facts.

Proposition 1. *For any tree $T = (V, E)$, the number of leaves is $\sum_{v \in N(T)} (deg_T(v) - 2) + 2$.*

Given an instance of the BDST problem with non-uniform degree bounds and a tree T, for node u in T, let $U_T(u)$ denote the *unused degree* of u in T, i.e., $U_T(u) = B_u - deg_T(u)$. For $V' \subseteq V$, the *unused degree* of vertex set V' of tree T, denoted by $U_T(V')$, is the summation of all unused degree of nodes in V'. The propositions consider unused degrees and the feasibility of BDST instances.

Proposition 2. *Let T' be a tree on V'. If for every node $u \in V'$, $deg_{T'}(u) \leq B_u$, we have that the unused degree $U_{T'}(V') = \sum_{u \in V'} (B_u - 2) + 2$.*

Proposition 3. *If $\sum_{u \in V} B_u < 2|V| - 2$, there is no solution to this instance.*

Corollary 2. *Let V_1 denote the set of nodes u such that $B_u = 1$. For any feasible instance, we have that $|V_1| \leq \sum_{w \in V - V_1} (B_w - 2) + 2$.*

Therefore, later on, we assume that the condition from Corollary 2 holds.

The next proposition states that there is a way to find a set of paths that spans the vertex set of a tree.

Proposition 4. *For any tree $T = (V, E)$ with k leaves, there exists a set of k edge-disjoint paths in T that spans T.*

Proof. (Sketch) Since T is rooted, we have a parent-child relationship in T. After each parent picks one of its children, the tree is decomposed into the required set of paths. $\square$

The next proposition, which can be proved using triangle inequality, summarizes useful properties related to Δ.

Proposition 5. *(1) For any v, $\ell(r, v) \leq \Delta$. (2) For any pair v and u, $\ell(u, v) \leq 2\Delta$.*

4.1.2 The Minimum Metric Traveling k-Salesperson Problem. Given an undirected complete graph $G = (V, E)$, a start node $r \in V$ and a nonnegative cost function for each edge e that satisfies triangle inequality, the *Minimum Metric Traveling k-Salesperson Problem* is to find a collection of k subtours, each containing the start node r, such that each node is in at least one subtour. The length of a k-tour is defined as the maximum length among its k subtours; an optimal k-tour is a k-tour of minimum length. Frederickson, Hecht, and Kim [2], presented an algorithm (later called the FHK algorithm) for the problem that guarantees $C_k \leq (e + 1 - 1/k)C_k^*$, where C_k is the cost of the maximum subtour returned by the algorithm, C_k^* is the cost of largest subtour in an optimal solution, and e is the best approximation ratio for the algorithm for approximating single TSP tour. Up to the current knowledge, $e = 1.5$ [7].

4.2. The base

We build a tree H from set of nodes in V_3. When all degree bounds are equal, Könemann, Levin, and Sinha present the following lemma.

Lemma 2. *[1, Theorem 1] Given V, a metric d over V, and maximum node-degree B, the Könemann* et al.*'s algorithm produces a tree with maximum node-degree B and diameter $O(\sqrt{\log_B n} \cdot \Delta)$ where Δ is the minimum diameter of any trees whose maximum degree is at most B.*

We present a simple reduction that suffices to give an approximation ratio of $O(\sqrt{\log n})$. Given an instance $\mathcal{I}$ of the non-uniform case, consisting of node set V, a metric d over V, and degree requirement B_v for each $v \in V$, we construct an instance $\mathcal{I}'$ of the uniform BDST problem as follows. For each $v \in V$, we create nodes $v^1, v^2, \dots v^{B_v - 2}$. Denote this set of nodes by $c(v)$. Let $V' = \cup_{v \in V} c(v)$ and define mapping $f : V' \to V$ to be such that $f(u) = v$ iff $u \in c(v)$. The metric d' can be defined naturally, i.e., for any

$u, v \in V'$, let $d'(u, v) = d(f(u), f(v))$. Note that, for any v, the distance between any two nodes in $c(v)$ under d' is zero. Finally, we assign the degree requirement for every node $v \in V'$ to 3.

The optimal solution to $\mathcal{I}$ can be easily converted to the solution to the BDST instance $\mathcal{I}'$ with the same diameter. Let tree T' denote the solution to $\mathcal{I}'$. The following lemma proves the converse.

Lemma 3. *The tree T' can be converted to tree T which is a solution to the original non-uniform problem whose diameter is at most the diameter of T'.*

Proof. For any graph H and subset U of H's nodes, let $H[U]$ denote an subgraph of H induced by U. We root T' at some arbitrary node $r \in T'$. We show that T' can be modified so that (1) for each $v \in V$, subgraph $T'[c(v)]$ are connected, and (2) the distance from any node u to r in T' does not increase.

We shall describe an operation on any node $v \in V$ that ensures that $T'[c(v)]$ satisfies condition (1) while keeping all distances satisfying condition (2).

Consider some node $v \in V$ such that $T'[c(v)]$ is not connected. Let A and B denote two connected components in $T'[c(v)]$. For any component C, define *root $R(C)$ of the component* to be the node closest to root r of T'. The *tails of component C*, denoted by $T(C)$ is a set of nodes in C, which are furthest from r. (When the component is a singleton, $R(C) \in T(C)$.)

Assume that $R(A)$ is no further from the root than $R(B)$. We shall join $R(B)$ to one of A's tails. Let p be parent of $R(B)$ in T'. Note that, p exists because the only case that $R(B)$ does not have a parent is when $R(B)$ is the root, but in that case, $R(B)$ should be closer to the root than $R(A)$, contradicting the assumption above.

Pick an arbitrary node $t \in T(A)$. To change $R(B)$'s parent to t, we first delete edge $(p, R(B))$ and add edge $(t, R(B))$. Note that, the distance from r to $R(B)$ is now equal to the distance from r to $R(A)$.

If degree of t is not violated, we are done. Otherwise, since $B_t = 3$, t now has 3 children including $R(B)$. Consider the children of t which are not $R(B)$; let u be one of them. Let t' be some node in $T(B)$. If t' has some child, w, remove edge (t', w) and add edge (p, w) so that t' has degree at most 2.

Note that, previously w before the operation w goes to p through a component in $T'[c(v)]$. Since edge (p, w) connects w to directly p; from the triangle inequality, the distance from w to r does not increase.

We shall change u's parent to t' by deleting edge (t, u) and adding edge (t', u). Note that, the distance from u to the root remains the same because t' and t are connected through a path of length zero and $d'(t, u) = d'(t', u)$.

We can repeat the operation until all $T'[c(v)]$ are connected for all $v \in V$. From this, T' can be converted to the solution T with the same diameter. $\square$

From this reduction and Lemma 2, we have the diameter bound of the base.

Lemma 4. *The diameter of the base is at most $O(\sqrt{\log n})\Delta$.*

4.3. The tails

We consider nodes in V_2. Let K be the total amount of unused degree in the base. We build K paths from V_2 and connect them to nodes in the base whose unused degrees are non-zero.

The following lemma states that for any tree that satisfies the degree constraints, the number of leaves is at most K, i.e., $\sum_{v \in V_3}(B_v - 2) + 2$.

Lemma 5. *Let $T = (V, E)$ be any tree. If for every node $u \in T$, $deg_T(u) \leq B_u$, the number of leaves is at most $K = \sum_{v \in V_3}(B_v - 2) + 2$.*

Proof. Let N denote the set of internal nodes in T. Using Proposition 1, we have that the number of leaves in T is $\sum_{v \in N}(deg_T(v) - 2) + 2$. From the degree constraints, we have that

$$\sum_{v \in N}(deg_T(v) - 2) + 2 \leq \sum_{v \in N}(B_v - 2) + 2 = \sum_{v \in V_3}(B_v - 2) + 2.$$

The last step follows because $N \subseteq V_2 \cup V_3$ and for any $v \in V_2$, $B_v - 2 = 0$. $\square$

Consider the optimal solution T^*. Let K' denote the number of leaves in T^*. From Lemma 5, we have that $K' \leq K$. We use Proposition 4 to partition nodes in T^* into K' paths. Clearly, the length of each path cannot be greater than Δ (because each path is a subpath of T^*). Note that, we can increase the number of such paths by cutting them into K paths without increasing each path's length. Therefore, we have the following lemma.

Lemma 6. *There exists a set of K edge-disjoint paths spanning nodes of T^* such that the length of each path is at most Δ.*

By adding edges from r to both ends of each path in the above lemma, we have a set of K tours spanning all nodes, each of length at most 3Δ. This is true because $\ell(r,u) \leq \Delta$, for any node u (see Proposition 5). Therefore, we have a feasible solution to the Minimum Metric Traveling k-Salesperson Problem with cost at most 3Δ.

We shall construct a set of K paths over nodes in V_2 whose lengths can be upper bounded by $3\alpha\Delta$, where α is the approximation guarantee of the FHK algorithm. Using the FHK algorithm, we partition V into K tours rooted at r. For each tour, we form a path by traversing the tour starting at the root node, and skipping over nodes in $V_3 - \{r\}$ and V_1, and finally removing r from the path. From the triangle inequality, removing nodes in $V_3 - \{r\}$ and V_1 does not increase the path length. Therefore, the cost of each path is at most $3\alpha\Delta$. Note that, each path contains only nodes in V_2. The following lemma states that the length of each path is not too long; its proof follows from the guarantee of the FHK algorithm and Lemma 6.

Lemma 7. *The length of each path is at most* $3\alpha\Delta$ *where* α *is the approximation guarantee of the FHK algorithm.*

By plugging in the approximation ratio of the FHK algorithm and Christofides' algorithm [7], we have the following corollary.

Corollary 3. *Each path is of length at most* $(15/2 - 3/K)\Delta$.

4.4. The main theorem

After building the base and the tails, we still need to attach each node in V_1 to the end of each tail. Proposition 5 implies that this would increase the diameter by at most Δ.

From the above discussion on V_1, using Lemma 4, Corollary 3, and Proposition 5, we have the main theorem.

Theorem 1. *There is a polynomial-time* $O(\sqrt{\log n})$*-approximation algorithm for the BDST with non-uniform degree bounds on* n *nodes.*

4.5. Removing the assumption on the degree bound of the root

We shall consider the case when every node v has $B_v \leq 2$. In this case, the optimal solution must be a path. Let Δ denote its diameter. If we connect two nodes of degree one in this path, we end up with a tour whose cost is at most 2Δ. Thus, we can use any β-approximation algorithm for the Traveling

Salesperson Problem (e.g., the Christofides algorithm [7]) to find a tour T whose cost is with in a factor of 2β of the optimal diameter Δ. (If we use the Christofides algorithm, the cost of T is at most 3Δ.)

We then need to find a tree that satisfies the degree bounds using T. Note that, the degree constraints are violated at nodes whose degree bound is one, but there can be at most two of them. To find the tree, we simply delete one edge from T incident to each node u whose $B_u = 1$ and add at most one edge to reconnect the path. This increases the cost of the tree by a factor of two. Thus, the diameter of the resulting tree is at most 6Δ.

Acknowledgement

We would like to thank anonymous reviewers for their very helpful comments.

References

[1] J. Könemann, A. Levin, A. Sinha, Approximating the degree-bounded minimum diameter spanning tree problem, in: Algorithmica, Springer, 2003, pp. 109–121.

[2] G. N. Frederickson, M. S. Hecht, C. E. Kim, Approximation algorithms for some routing problems, SIAM Journal on Computing 7 (2) (1978) 178–193.

[3] D. Ren, Y.-T. Li, S.-H. Chan, On reducing mesh delay for peer-to-peer live streaming, in: INFOCOM, 2008, pp. 1058–1066.

[4] R. Ravi, Rapid rumor ramification: approximating the minimum broadcast time, in: FOCS, 1994, pp. 202–213.

[5] M. Elkin, G. Kortsarz, Sublogarithmic approximation for telephone multicast, Journal of Computer and System Sciences 72 (4) (2006) 648 – 659.

[6] F. Huang, B. Ravindran, A. V. Kumar, An approximation algorithm for minimum-delay peer-to-peer streaming, in: IEEE Ninth International Conference on Peer-to-Peer Computing., 2009, pp. 71–80.

[7] N. Christofides, Worst-case analysis of a new heuristic for the traveling salesman problem, in: Symposium on New Directions and Recent Results in Algorithms and Complexity, 1976, p. 441.