

A Flow Formulation for the Optimum Communication Spanning Tree

Elena Fernández, Carlos Luna-Mota^{1,2}

*Department of Statistics and Operations Research
Universitat Politècnica de Catalunya - Barcelona Tech
Barcelona, Spain*

Achim Hildenbrandt, Gerhard Reinelt, Stefan Wiesberg^{3,4,5}

*Institut für Informatik
Universität Heidelberg
Heidelberg, Germany*

Abstract

In this paper we address the Optimum Communication Spanning Tree Problem. We present a formulation that uses three index variables and we propose several families of inequalities, which can be used to reinforce the formulation. Preliminary computational experiments are very promising.

Keywords: Spanning tree, Optimum Communication Spanning Tree.

¹ Email: `e.fernandez@upc.edu`

² Email: `carlos.luna-mota@upc.edu`

³ Email: `achim.hildenbrandt@informatik.uni-heidelberg.de`

⁴ Email: `gerhard.reinelt@informatik.uni-heidelberg.de`

⁵ Email: `stefan.wiesberg@informatik.uni-heidelberg.de`

1 Introduction

In the *Optimum Communication Spanning Tree Problem* (OCSTP) a given set of n origin/demand points with communication requirements between them must be connected by a tree. The communication cost between an origin/destination pair, depends both on the communication requirement and on the distance on the tree between the nodes of the pair. The OCSTP is to find a spanning tree of minimum total communication cost.

The OCSTP was originally introduced by Hu [9]. Despite of its apparent simplicity, the OCSTP turns out to be a very challenging combinatorial optimization problem. Johnson, Lenstra, and Rinnooy Kan [10] proved that finding an OCST is $\mathcal{NP}$ -hard, even if all origin/destination pairs have the same communication requirement. Moreover, unless $\mathcal{NP} = \mathcal{P}$, no polynomial time approximation scheme exists since the problem is $\mathcal{MAX SNP}$ -hard [12]. Yet, [9] presented two particular cases of the OCSTP defined on a complete graph that can be solved in polynomial time. The first one is the Optimum Requirement Spanning Tree Problem where, in the original graph, the distance between all pairs of nodes is the same. In this case, an optimal solution is given by a Gomory-Hu tree of an associated network. The second one is the Optimum Distance Spanning Tree Problem, where all pairs of nodes have the same communication requirements. Now, under some additional assumptions, the optimal solution has a star topology.

In addition to a broad range of applications within telecommunication, transportation, and computer networks, the OCST appears as a subproblem in some Network Hub Location Problems. For instance, in the Tree-of-Hubs Location Problem (THLP) (see [5,6]), when both the set of hubs and the allocation pattern of nodes to hubs are known, the problem can be polynomially transformed into an OCSTP. Thus, efficient solution procedures for the OCSTP may also serve as subroutines for solution procedures for the THLP.

To the best of our knowledge, there is only one algorithm, due to Ahuja and Murty [1], for optimally solving the general case of the OCSTP. It is a combinatorial branch & bound algorithm, which uses lower planes to obtain a valid lower bound at every node of the branch & bound tree. With this algorithm the authors were able to prove the optimality of the obtained solutions for sparse instances with up to 40 nodes. Rothlauf [13] modeled the OCSTP as an integer linear program, which was able to obtain optimal solutions for instances with up to 12 nodes in reasonable computational times. Contreras [2] (see also [3]) proposed another integer programming formulation which, enforced with additional valid inequalities, was able to produce optimal so-

lutions for instances with up to 25 nodes in reasonable computational times. Contreras, Fernández and Marín [4] presented a Lagrangean relaxation which produced good lower and upper bounds for instances with up to 50 nodes. To the best of our knowledge no exact algorithm exists able to solve medium sized instances for the general case.

One of the main issues in the search for a successful formulation for the OCSTP is to find a trade off between the number of variables in the formulation and the tightness of the Linear Programming (LP) bound. It is known that formulations with path-based 4 index variables produce very tight lower bounds for the OCSTP [4]. However, the main drawback of such formulations is the number of variables they require, since each origin/demand pair requires as many variables as arcs the network has. In practice, this makes it impossible to solve instances with more than 30 nodes with a general purpose solver. On the other hand, it is possible to formulate the OCSTP using only 2 index variables [7] but, for the moment, such formulations are not tight enough so as to solve instances with more than 25 nodes in reasonable computing times. In this paper we study formulations for the OCSTP with 3 index variables, which can be seen as a compromise between the above two types of formulations. For each fixed origin, the 3 index variables are obtained by aggregating over all possible destinations the path-based 4 index variables. The resulting formulation models the intersection of a series of spanning trees, each of them rooted at one different origin, in the spirit of recent formulations for other problems [8,11]. Basic versions of 3 index formulations are still incapable of solving moderate size instances. Nevertheless, preliminary results indicate that they can be reinforced with different types of valid inequalities so as to yield promising results.

The structure of the paper is the following. In Section 2 we first recall the formal definition of the Optimum Communication Spanning Tree Problem. Then, we introduce the basic 3 index formulation for the OCSTP with which we will work. Section 3 presents different types of valid inequalities, which reinforce the basic formulation. In Section 4 we conclude the paper with some comments and guidelines for further research.

2 Optimum Communication Spanning Trees

The Optimum Communication Spanning Tree Problem (**OCSTP**) is formally defined as follows. Let $G=(V, E)$ be a complete undirected graph with $n = |V|$ nodes and $m = |E|$ edges, $c : E \rightarrow \mathbb{R}^+$ a cost function over the edges, and $r : V \times V \rightarrow \mathbb{R}^+ \cup \{0\}$ a communication requirement function between pairs of

nodes. Without loss of generality we assume that $r_{ji} = 0$ for $j > i$. If needed we redefine r_{ij} as $r_{ij} + r_{ji}$ for all $i, j \in V$, $i < j$.

The solution space of the OCSTP is the set of all spanning trees of G . For a given spanning tree T of G ,

the communication cost over T of a pair of nodes $i, j \in V$, $i \neq j$, is defined as the cost (length) with respect to c of the (unique) path in T that connects i and j , c_{ij}^T , multiplied by the communication requirement between the pair: $c_{ij}^T r_{ij}$.

the total communication cost of T is the sum of the communication costs over T of all pairs of nodes: $\sum_{i \neq j} c_{ij}^T r_{ij}$. Note that the total communication cost of T can be evaluated in $O(n^2)$ time.

The OCSTP is to find a spanning tree of G of minimum total communication cost.

Remark 2.1 There is an alternative way of computing the communication cost of a tree T , which takes into account that T contains a unique *communication path* connecting each origin/destination pair. Let f_{ij}^T denote the total amount of flow that circulates through any of the two directions of (i, j) . Note that f_{ij}^T is the sum of the communication requirements of all the origin/destination pairs that use either (i, j) or (j, i) in their communication path. Then, the communication cost of T can also be computed as $\sum_{i, j \in V: j > i} c_{ij} f_{ij}^T$.

To formulate the OCST we use a set of binary variables which indicate the edges that are used in the tree. For all $i, j \in V$, $i < j$, let $x_{ij} = 1$ if edge (i, j) is in the tree, and 0 otherwise. In addition, for each $u \in V$ we define a continuous set of variables, denoted f_{uij} , to indicate how the flow $O_u = \sum_{v > u} r_{uv}$ originated at u circulates through the tree defined by x . For $u, i, j \in V$, $i \neq j$, f_{uij} indicates the flow with origin in u through arc (i, j) . To represent the arcs that are used for sending the flow originated at u , we define for all $i, j \in V$, $i \neq j$, a binary decision variable y_{uij} which takes the value 1 if arc (i, j) is used for sending the flow originated at u . Then, a valid formulation for the OCST is:

$$\begin{aligned}
(1) \quad & \min \sum_{i, j \in V: i < j} c_{ij} \sum_{u \in V} (f_{uij} + f_{uji}) \\
(2) \quad & \sum_{j \in V \setminus \{u\}} f_{uoj} = O_u \quad u \in V \\
(3) \quad & \sum_{j \in V \setminus \{i, u\}} f_{uij} - \sum_{j \in V: j \neq i, j \neq u} f_{uji} = -r_{ui} \quad u, i \in V
\end{aligned}$$

$$\begin{aligned}
(4) \quad & f_{uij} \leq M y_{uij} \quad u, i, j \in V, j > i \\
(5) \quad & f_{uji} \leq M y_{uji} \quad u, i, j \in V, j > i \\
(6) \quad & \sum_{i,j \in V; j \neq i, j \neq u} y_{uij} = n - 1 \quad u \in V \\
(7) \quad & y_{uij} + y_{uji} \leq x_{ij} \quad u, i, j \in V, j > i \\
(8) \quad & \sum_{i,j \in V; j > i} x_{ij} = n - 1 \\
(9) \quad & x_{ij}, y_{uij} \in \{0, 1\}, f_{uij} \geq 0 \quad i, j, u \in V, j > i
\end{aligned}$$

For each $u \in V$, constraints (2)–(3) model a flow of value O_u with origin at u and destination r_{uv} for $v \in V$. Constraints (4) and (5) relate the f and the y variables by activating the arcs that are used for routing the flows. Together with cardinality constraints (6), constraints (2)–(5) guarantee that, for each $u \in V$, the arcs used for sending the flow O_u define a spanning tree. Constraints (7) relate the x and the y variables by imposing that if an arc is used for sending some flow, then associated edge is in the tree. These constraints will hold as equality for instances where communication exists between all pairs of nodes, i.e. $r_{ij} + r_{ji} > 0$ for all $i, j \in V, j > i$. Otherwise, some of these constraints will hold as strict inequality. Finally, constraint (8) guarantees that all the flows define the same spanning tree, by limiting the total number of edges to $n - 1$.

In constraints (4) and (5) M is a sufficiently large constant, whose value may affect significantly the effectiveness of the constraints. We use $M = O_u - rm_u$, where $rm_u = \min_{v>u} r_{uv}$.

Indeed, there are several alternatives for modeling the relation between the f and x variables. In our formulation, this is done via the intermediate y variables and constraints (4)–(5) and (7). An alternative would be to omit the y variables and to substitute the above three sets of constraints by

$$(10) \quad f_{uij} + f_{uji} \leq M x_{ij} \quad u, i, j \in V, j > i.$$

However, for instances where communication requirements do not exist between all pairs of nodes, our formulation may give tighter LP bounds because, as mentioned, some inequalities (7) will hold as strict inequality.

3 Valid inequalities

In this section we present several families of valid inequalities that can be used to reinforce formulation (1)–(9).

(a) **Vertex cutset inequalities.** For all $u \in V$ at least one arc must leave

vertex u , i.e.

$$\sum_{j \in V \setminus \{u\}} y_{uj} \geq 1.$$

On the other hand, for all $u, i \in V$, $u \neq i$ exactly one arc associated with the flow emanating from u must enter i , i.e.

$$\sum_{j \in V \setminus \{i\}} y_{uj} = 1.$$

(b) **Set cutset inequalities.** For $S \subset V$

$$x(\delta(S)) \geq 1.$$

(c) **Min-cut values inequalities.** Let m_{ij} denote the value of the min-cut separating i and j in the original graph, relative to a capacity vector given by r . For all $i, j \in V$, $i < j$, the following inequalities are valid:

$$m_{ij}x_{ij} \leq \sum_{u \in V \setminus \{j\}} f_{uij} + \sum_{u \in V \setminus \{i\}} f_{uji}.$$

$$\sum_{u \in V \setminus \{j\}} r_{uj}y_{uij} + \sum_{u \in V \setminus \{i\}} r_{ui}y_{uji} \leq \sum_{u \in V \setminus \{j\}} f_{uij} + \sum_{u \in V \setminus \{i\}} f_{uji}.$$

(d) **Minimum flows trough arcs.** For $u, i, j \in V$, $i < j$, the following inequality must hold:

$$r_{uj}y_{uij} + r_{ui}y_{uji} \leq f_{uij} + f_{uji}.$$

Let $k_1 = \max\{m_{uj}, m_{ij}\}$ and $k_2 = \max\{m_{iu}, m_{ij}\}$. The following lower bound on the flows must hold:

$$k_1y_{uij} + k_2y_{uji} \leq \sum_{u \in V \setminus \{j\}} f_{uij} + \sum_{u \in V \setminus \{i\}} f_{uji}.$$

4 Comments and guidelines for further research

We have run a series of preliminary experiments with benchmark instances from the literature using a general purpose commercial solver.

Broadly speaking the obtained numerical results indicate the following:

- The bound of the LP relaxation of formulation (1)–(9) lies within 60–70% of the optimal value. As could be expected, the quality of the bound seems to deteriorate as the size of the instances increase.
- The effect of the cutset inequalities presented in items (a) and (b) seem to be very limited. For the moment we have no indication that they can be useful for the proposed formulation.
- The inequalities which seem to have more impact are the ones of item (c), derived from min-cut values. Depending on the instance, these inequalities reinforce the lower bound to 80–95% of the optimal value.
- The inequalities imposing lower bounds on the flows through used arcs, presented in item (d) also improve the quality of the lower bound. This additional improvement is larger in the instances for which the min-cut values inequalities were not so effective. When both families are added the obtained lower bounds are beyond 90% of the optimal values.
- The computing times can be improved by removing constraints (b) and (c) from the original formulation and by dynamically separating them. That is, they are incorporated to the formulation only if they are violated by the current LP solution.

From the results obtained so far it is clear that further research is needed to assess the interest of the proposed formulation. The obtained results suggest exploring further types of inequalities derived from min-cut values. Another avenue of research focuses on obtaining tight bounds for the costs of the flows through the used origin/destination paths.

Acknowledgement

This research has been partially funded through joint grants AIB2010DE-00137 by the Spanish Ministry of Science and Education, and DDAD50749416 by the German Academic Exchange Service in the Project-based Personnel Exchange Program. The research of E. Fernández and C. Luna-Mota has been partially funded through grant MTM2009–14039–C06–05 of the Spanish Ministry of Science and Education and ERDF funds.

References

- [1] Ahuja, R.K., and V.V.S. Murty, *Exact and Heuristic Algorithms for the Optimum Communication Spanning Tree Problem*, *Transportation Science* **21**

(1987), 163–170.

- [2] Contreras, I., “Network Hub Location – Models, Algorithms, and Related Problems”, Ph.D. thesis, Universitat Politècnica de Catalunya, 2009.
- [3] Contreras, I., and E. Fernández, *General Network Design: A Unified View of Combined Location and Network Design Problems*, European Journal of Operational Research **219** (2012), 680–697.
- [4] Contreras, I. and E. Fernández, and A. Marín, *Lagrangian bounds for the optimum communication spanning tree problem*, TOP **18** (2010), 140–157.
- [5] Contreras, I. and E. Fernández, and A. Marín, *Tight bounds from a path based formulation for the Tree-of-Hubs Location Problem*, Computers & Operations Research **36** (2010), 3117–3127.
- [6] Contreras, I. and E. Fernández, and A. Marín, *The tree-of-hubs location problem: A comparison of formulations*, European Journal of Operational Research (202), 390–400.
- [7] Fernández, E., and C. Luna-Mota, and G. Reinelt, *A compact formulation for the optimum communication spanning tree problem*, International Symposium on Mathematical Programming (ISMP2012), Berlin, 2012.
- [8] Gouveia, L. and J. Telhada, *Reformulation by Intersection Method on the MST Problem with Lower Bound on the Number of Leaves*, Lecture Notes in Computer Science, Proceedings of INOC 2011 (2011)**6701** 83–91.
- [9] Hu, T.C., *Optimum Communication Spanning Trees*, SIAM J. Comput. **3** 1974,(188–195).
- [10] Johnson, D.S., and J.K. Lenstra, and A.H.G. Rinnooy Kan, *The Complexity of the Network Design Problem*, Networks **8** (1978), 279–285.
- [11] Martinez, L., and A. Salles da Cunha, *The min-degree constrained minimum spanning tree problem: Formulations and Branch-and-cut algorithm*, Discrete Applied Mathematics (special issue dedicated to ISCO 2010), to appear.
- [12] Papadimitriou, C., and M. Yannakakis, *Optimization, approximation, and complexity classes*, STOC 88: Proceedings of the twentieth annual ACM symposium on theory of computing (1988), 229–234.
- [13] Rothlauf, F., *On Optimal Solutions for the Optimal Communication Spanning Tree Problem*, Operations Research **57** (2009), 413–425.