Constrained Space Information Flow

Alfred Uwitonze[1], Jiaqing Huang[1(✉)], Yuanqing Ye[2], and Wenqing Cheng[1]

[1] School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, People's Republic of China
jqhuang@mail.hust.edu.cn

[2] Department of Electrical and Computer Engineering, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA

Abstract. *Space Information Flow* (SIF), also known as Space Network Coding, is a new research paradigm which studies network coding in Euclidean space, and it is different with *Network Information Flow* proposed by Ahlswede *et al*. This paper focuses on the problem of *Constrained Space Information Flow* (CSIF), which aims to find a min-cost multicast network in 2-D Euclidean space under the constraint on the number of relay nodes to be used. We propose a new *polynomial*-time heuristic algorithm that combines Delaunay triangulation and linear programming techniques to solve the problem. Delaunay triangulation is used to generate several candidate relay nodes, after which linear programming is applied to choose the optimal relay nodes and to compute their connection links with the terminal nodes. The simulation results shows the effectiveness of the proposed algorithm.

Keywords: Network Information Flow · Delaunay triangulation · Network coding in space · Space Information Flow

1 Introduction

Departing from *Network Information Flow* (NIF) proposed by Ahlswede *et al.* [1] in 2000, *Space Information Flow* (SIF) [2,3] is a new concept proposed by Li and Wu in 2011 and it studies *network coding in Euclidean space*. SIF is also different with both Euclidean Steiner Minimal Tree (ESMT) [4] and Minimum Spanning Tree (MST) [5]. ESMT is the optimal routing in space. MST connects together all the terminals of a given set with a shortest network, without any additional relay node, while additional relay nodes are required in SIF [2,3]. The pentagram [6] example illustrated in Fig. 1 demonstrates that SIF can strictly outperforms ESMT, with the *cost advantage* [7] being strictly bigger than 1. The *cost advantage* is defined as the ratio of the minimum network cost without network coding over that with network coding. Consider six multicast terminal nodes in a 2-D Euclidean space depicted in Fig. 1(a). Among the six multicast terminals, five (T_1 to T_5) are equally placed on a circle and form a regular pentagon whose center is node O. The circumscribed circle of the pentagon has a radius of 1. Node O is selected as the multicast source, while the

Q. Chen et al. (Eds.): ChinaCom 2016, Part II, LNICST 210, pp. 475–485, 2018.
DOI: 10.1007/978-3-319-66628-0_45

remaining five nodes (T_1 to T_5) are the receivers. With ESMT, an optimal solution can be computed [8] and the cost is 4.6400/bit (Fig. 1(b)). Three Steiner nodes (S_1 to S_3) are introduced for connecting the terminal nodes, each adjacent to three links which form three angles of 120°. An optimal solution by SIF is depicted in Fig. 1(c). The total distance is 9.1354, while every sink receives 2 bits. The normalized cost is $9.1354/2 = 4.5677$/bit. Five relay nodes (R_1 to R_5) are introduced for connecting the terminal nodes, each adjacent to three links which form three angles of 120°. The *cost advantage* of the pentagram example is $4.6400/4.5677 \approx 1.0158 > 1$. Despite its small value, we emphasize that the gap between the two optimal costs reveals that multicast with SIF is fundamentally a different problem from geometric ESMT, with a different problem structure, and probably a different computational complexity. The placement of relay nodes in wireless sensor networks is a potential application of SIF [9].

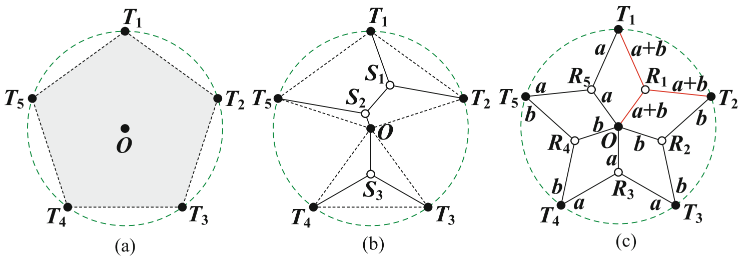

Fig. 1. Illustration example of *pentagram*. (a) Six terminal nodes in 2-D Euclidean space; (b) Optimal solution with ESMT (cost = 4.6400/bit); (c) Optimal solution with SIF (cost = 4.5677/bit).

For SIF, Li and Wu [3] studied the problem of multiple-unicast network coding in space. Yin *et al*. [10] proved a number of properties of optimal multicast network coding in 2-D Euclidean space. Xiahou *et al*. [11] applied SIF as a tool to design a framework for analyzing the network coding conjecture. A heuristic approach based on iterative method has been proposed by Hu *et al*. [12] to address min-cost video multicast problem via Constrained SIF. A polynomial-time heuristic algorithm for computing the optimal SIF solution in multicast network has been proposed by Huang *et al*. [6]. In a subsequent study, Huang and Li [9] presented a polynomial-time heuristic approach based on non-uniform recursive space partitioning for computing SIF. In another subsequent work, Uwitonze *et al*. [13] presented a polynomial-time heuristic approach based on Delaunay triangulation that computes the SIF solutions in multicast networks. In line with routing in space, Gilbert and Pollack [4] studied the properties of optimal ESMT. As for MST, its complexity is polynomial [5].

The objective of SIF is to minimize the cost of constructing a network, allowing network coding to be used and additional relay nodes to be inserted for connecting a given set of terminals in geometric space, while satisfying end-to-end throughput demands among terminals [2]. However, adding more relay nodes may clearly lead to a higher cost in practice, given that each extra relay node may be associated with hardware and deployment cost. Therefore, it is necessary to consider such cost by minimizing the number of additional relay nodes. In this paper, we propose the Constrained SIF (CSIF) problem, which is a new version of SIF that considers the transmission of information flows in a geometric space under the constraint (restriction) on the number of additional relay nodes that can be introduced to connect a set of given terminal nodes. The space we consider in this work is a 2-D Euclidean space. To the best of our knowledge, this is the first work to explore the problem of Constrained SIF (CSIF) and to use Delaunay Triangulation (DT) [14] in CSIF. DT has two properties that are useful to reduce the overall length of the tree, as denoted by Smith *et al.* [14]. Firstly, since MST of N is contained in the $DT(N)$, a number of edges in ESMT is the same as edges in MST. Secondly, since each Delaunay triangle tends to be equilateral, we achieve the maximum possible reduction in using the ESMT, as compared with using the MST.

The main contribution of our paper can be summarized as follows:
We propose the first heuristic algorithm based on Delaunay Triangulation (DT) and Linear Programming (LP) techniques, with a *polynomial*-time complexity that computes the min-cost in multicast networks and the corresponding network topology (including the way relay nodes are connected with the terminal nodes, as well as the flow rate on the connection links), under the constraint on the number of additional relay nodes to be introduced.

The rest of this paper is organized as follows: Sect. 2 discusses the problem formulation. Section 3 describes the detailed steps of the new heuristic algorithm for CSIF. Section 4 presents the simulation results, while Sect. 5 concludes the paper.

2 Problem Formulation

This work focuses on the problem of min-cost multicast network coding in 2-D Euclidean space. For $N \geq 3$ given terminal nodes T_1, T_2,. . . ,T_N in the Euclidean space and a multicast session from one source to a number of sinks, the objective is to compute a min-cost multicast transmission scheme using SIF, that permits to insert at most M extra relay nodes. The total network cost is defined as $\sum_{uv} w(uv)f(uv)$, where $f(uv)$ denotes the information flow rate on a link uv in space, while $w(uv)$ denotes the weight of the link uv, and it is equal to the Euclidean distance $||uv||$ of uv [2,3]. These two variables are called *positions* and *flow assignments*. The connection topology of all nodes will be determined by flow assignments, because a link with a zero rate shows that the link does not exist. Our goal is to achieve the min-cost by tuning these two sets of variables with no more than M relay nodes.

3 The Proposed Heuristic Algorithm for CSIF

3.1 The Main Idea of Heuristic Algorithm

The main idea of our algorithm is to use at most M relay nodes to establish a min-cost multicast network connection from $N \geq 3$ given terminal nodes in space. Before introducing the constraint number of relay nodes M, the algorithm uses two alternative strategies to retain the relay nodes from LP computation: 1DT-2DT strategy and 2DT-1DT strategy. With 1DT-2DT strategy, the algorithm retains the less possible candidate relay nodes first, followed by retaining the most possible candidate relay nodes. With 2DT-1DT strategy, the algorithm retains the most possible candidate relay nodes first, followed by retaining the less possible candidate relay nodes. The less possible candidate relay nodes here refer to the candidate relay nodes generated in triangles, while the most possible candidate relay nodes refer to the candidate relay nodes generated in quadrilaterals. Quadrilaterals are obtained by concatenating two adjacent Delaunay triangles.

3.2 Detailed Description of Heuristic Algorithm

The proposed algorithm is based on DT and LP techniques. DT is used for generating at most $(2N-5)$ Delaunay triangles from $N \geq 3$ given terminal nodes [14]. Subsequently, it helps to compute a number of candidate relay nodes from all Delaunay triangles and quadrilaterals. LP is applied to choose the optimal relay nodes and to compute their connection links with the terminals. The proposed algorithm adopts the following LP model:

Minimize $cost = \sum_{\overrightarrow{uv} \in A} w(\overrightarrow{uv}) f(\overrightarrow{uv})$
Subject to :

$$\begin{cases} \sum_{v \in V_{\uparrow}(u)} f_i(\overrightarrow{vu}) = \sum_{v \in V_{\downarrow}(u)} f_i(\overrightarrow{uv}) & \forall i, \forall u \\ f_i(\overrightarrow{T_i S}) = r & \forall i \\ f_i(\overrightarrow{uv}) \leq f(\overrightarrow{uv}) & \forall i, \forall \overrightarrow{uv} \\ f(\overrightarrow{uv}) \geq 0, f_i(\overrightarrow{uv}) \geq 0 & \forall i, \forall \overrightarrow{uv} \end{cases} \tag{1}$$

The LP model (Eq. (1)) is based on undirected network $G = (V, E)$, where $V = N \cup R$, N denotes the set of terminal nodes and R is the set of extra relay nodes, while E denotes the set of undirected links. There are bi-directed possibilities of transmission in space. Therefore, we make links bi-directed and denote a set of directed links as $A = \{\boldsymbol{uv}, \boldsymbol{vu} | uv \in E\}$. In the objective function, the decision variable $f(\overrightarrow{uv})$ is regarded as the combined effective flow rate on a link $\overrightarrow{uv}$. The coefficient $w(\overrightarrow{uv})$ equals to the Euclidean distance $|\overrightarrow{uv}|(=|\overrightarrow{vu}|=|uv|)$. In the LP constraints, $f_i(uv)$ is regarded as the rate of information flow from the source S to sink T_i on a link $\overrightarrow{uv}$. Such kinds of information flow are *conceptual* because they share instead of competing for available bandwidth on the same link. $f(\overrightarrow{uv})$ of a link uv equals to the maximum among all $f_i(\overrightarrow{uv})$. The constraint $\sum_{v \in V_{\uparrow}(u)} f_i(\overrightarrow{vu}) = \sum_{v \in V_{\downarrow}(u)} f_i(\overrightarrow{uv})$ guarantees the conceptional flow equilibrium property for every node and every conceptual flow i. We have both $f_i(\overrightarrow{uv})$ and $f_i(\overrightarrow{vu})$ to indicate the flows in two directions. $V_{\uparrow}(u)$ and $V_{\downarrow}(u)$ respectively denote upstream and downstream adjacent set of u in V. The constraint $f_i(\overrightarrow{T_iS}) = r$ characterizes the desired receiving rate at each terminal. The constraints $f(\overrightarrow{uv}) \geq 0$ and $f_i(\overrightarrow{uv}) \geq 0$ give the trivial bound. The detailed steps of the algorithm are shown in Algorithm 1.

Algorithm 1. A Heuristic Algorithm for CSIF

Require: Input: N ($N \geq 3$) terminal nodes, a multicast session
Ensure: Output: a CSIF solution
Initialize the total set of candidate relay nodes $R_{total} = \emptyset$;
Construct all the DT-triangles by Delaunay triangulation;
Initialize the subset of candidate relay nodes $R(x) = \emptyset$, $R'(x) = \emptyset$, MINCOST=$+\infty$;
for $x = 1$ *to* 2, **do**
  Construct polygons P_i of 3 and 4 edges by concatenating x adjacent DT-triangles;
  Construct the MST of each polygon P_i;
  Obtain the candidate relay nodes $R(x)$;
  Construct a complete graph with $(N + \sum_{x=1}^{x} |R(x)|)$ nodes;
  Solve the LP model based on the complete graph and output MINCOST and the corresponding resulting relay nodes $R'(x)$;
  if There are k ($k \geq 2$) adjacent relay nodes only **then**
    Use 2DT-1DT strategy to retain $R'(x)$ from polygon P_i of 4 edges only
  else
    Use 1DT-2DT strategy to retain $R'(x)$ from polygon P_i of 3 edges only
  end if
end for
Calculate $R_{total} = \bigcup_{x=1}^{2} R'(x)$;
for $R_{total} = 1$ *to* M, **do**
  Construct a complete graph with $(N + M)$ nodes;
  Solve the LP model based on the complete graph and output the CSIF *cost*;
  if *cost* < MINCOST **then**
    MINCOST=*cost*
  end if
  Compute the ESMTs of each polygon P_i;
  Place all ESMTs on a hierarchical priority queue Q based on the value of $\Delta = \frac{MST(P_i) - ESMT(P_i)}{ESMT(P_i)}$;
  Construct the network topology by picking ESMTs from Q in the same way as the Kruskal's algorithm;
end for
if The flow rates of all constrained relay nodes == 0 **then**
  Output MINCOST and stop.
end if

3.3 Complexity Analysis of Our Algorithm

Our heuristic algorithm considers all the possible candidate Steiner nodes generated from all Delaunay triangles concatenations as possible candidate relay nodes. According to [15], the method to obtain the Steiner nodes in a triangle and a quadrilateral is shown in Figs. 2 and 3, respectively.

Triangle: As depicted by Fig. 2, assume $\angle VUW$ is the biggest angle in $\triangle UVW$. If $\angle VUW \geq 120°$, then the Steiner node S is the vertex U. If $\angle VUW < 120°$, draw two equilateral triangles on any of the two edges of $\triangle UVW$, e.g., $\triangle UVX$ and $\triangle UWY$, then the Steiner node S is the intersection of the lines VY and WX. Thus, the time complexity is polynomial.

Quadrilateral: As stated by [15], the process to obtain the Steiner nodes in a convex quadrilateral $UVWX$ consists of three steps, as illustrated in Fig. 3. First, draw two equilateral triangles $\triangle UVY$ and $\triangle WXZ$. Next, draw two circles which pass at the vertices of the two equilateral triangles $\triangle UVY$ and $\triangle WXZ$. Last, draw the line YZ and the two steiner nodes S_1 and S_2 are the intersection of the line YZ with the two circles, as depicted in Fig. 3. Hence, the time complexity is also polynomial.

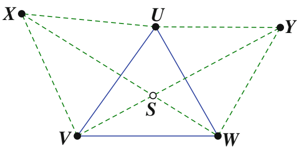

Fig. 2. Computing the candidate Steiner node in a triangle.

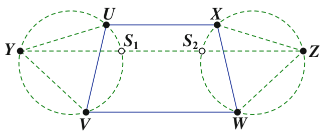

Fig. 3. Computing the candidate Steiner nodes in a quadrilateral.

The time complexity of DT is $O(N \log N)$ [16]. The time complexity of computing the candidate Steiner nodes for every Delaunay triangle and quadrilateral formed by concatenating two neighboring Delaunay triangles is polynomial. Given $N \geq 3$ terminal nodes, we can get at most $(2N-5)$ Delaunay triangles and $(3N-6)$ edges by DT. It is possible to concatenate at most $(N-2)$ neighboring triangles and $2N-5$ quadrilaterals, respectively. Hence, $|R_{total}| \leq 6N-16$, and the time complexity of LP is $O((N+|R_{total}|)^2) = O((7N-16)^2) = O(N^2)$. Thus, the time complexity of our algorithm is $O(N^3 \log N)$, which is polynomial.

4 Simulation Results

We have simulated our heuristic algorithm in 2-D Euclidean space. Our simulations used MATLAB to solve LPs. In multicast networks, the number of relay nodes required for an optimal solution is upper-bounded by $(N-2)$ for $h = 1$

and $(2N-3)(2N-2)$ for $h=2$ [10], where N is the number of the terminals and h is the multicast throughput. Thus, we set $M<(N-2)$ for $h=1$ and $M<(2N-3)(2N-2)$ for $h=2$, where M is the constraint on the number of relay nodes. All the tested cases correspond to $h=1$, except the pentagram network, which corresponds to $h=2$ [6]. The optimal ESMT is computed by GeoSteiner 3.1 that implements an exact ESMT algorithm [8]. The MST is computed by implementing Prim's algorithm [5] in MATLAB. For each tested case, one node is set as the source, while the remaining nodes are set as the terminals.

4.1 Cases of 10 Nodes Data Sets from OR-Library

We applied our algorithm to 10-points ($N=10$) data sets from OR-Library [17], which contained 15 cases with different positions. We set $M<8$ and we evaluated the performance of our algorithm by comparing the results of CSIF with SIF, optimal ESMT and MST. We define the gap between SIF and CSIF as $gap = \frac{SIF}{CSIF}$. Figure 4 shows the MST cost, min-cost for CSIF, SIF cost and optimal ESMT cost for all the 15 cases. Both SIF and optimal ESMT achieve the same results, since SIF degrades into optimal ESMT when $h=1$ [10]. CSIF outperforms MST for all the cases. Moreover, CSIF cost is very close to both SIF and ESMT costs for all the 15 cases (See Fig. 4). Table 1 shows the gap for all the 15 cases. The gap ≈1 for almost all the cases.

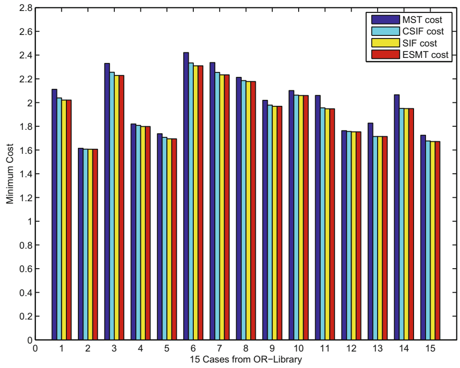

Fig. 4. MST cost, min-cost for CSIF when $M<8$, SIF cost and ESMT cost.

4.2 The Pentagram Network

We applied our algorithm to the pentagram network, where $N=6$. We set $M=5$ and the obtained CSIF topology is shown in Fig. 5. Figure 6 shows the optimal topology with SIF. Both SIF and CSIF achieve the same results (cost = 4.5677/bit), bacause they both use the same number of relay nodes, as

Table 1. Gap $= \frac{SIF}{CSIF}$

Case	Gap	Case	Gap	Case	Gap
1	0.9910367784	6	0.9897108913	11	0.9957873959
2	0.9997107509	7	0.9909349461	12	0.9979368571
3	0.9876711421	8	0.9964208849	13	0.9999795101
4	0.9949909489	9	0.9949003406	14	0.9993935543
5	0.9928556449	10	0.9982077096	15	0.9970678120

it can be seen in Figs. 5 and 6 and the gap = 1. Hence, the algorithm achieves the optimal solutions for the pentagram network. Figures 7 and 8 show the MST (cost = 5.0000/bit) and ESMT (cost = 4.6421/bit) topologies, respectively. CSIF outperforms both MST and ESMT in terms of min-cost.

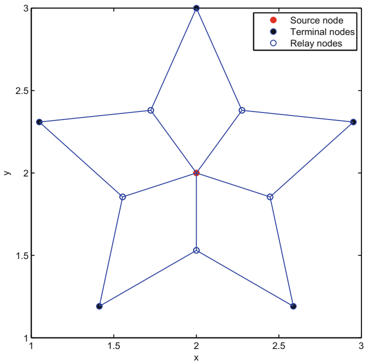

Fig. 5. CSIF result for Pentagram network.

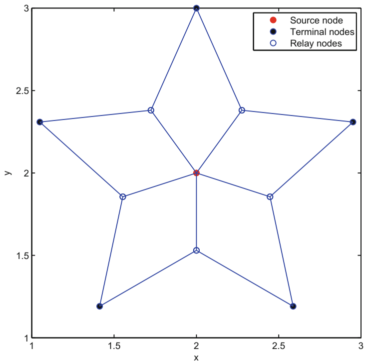

Fig. 6. SIF result for Pentagram network.

4.3 Random Networks

We tested the algorithm in random networks, which are generated by the Waxman model [18]. Throughout our simulations, we observed that in most of the cases for such random networks, gap ≈1 when $M < (N - 2)$. Furthermore, CSIF outperforms MST for all tested cases. Figure 9 illustrates the CSIF result (cost = 1.5776/bit) for one example of such cases when $N = 8$ and $M = 2$. Figure 10 shows the SIF result (cost = 1.5771/bit), Fig. 11 shows the MST result (cost = 1.6053/bit), while Fig. 12 shows the optimal ESMT result (cost = 1.5771/bit). Both SIF and optimal ESMT achieve the same results, since SIF degrades into optimal ESMT when $h = 1$ [10]. The gap = 0.9996.

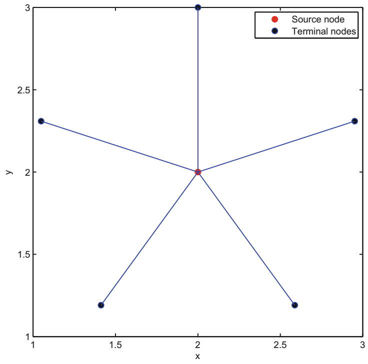

Fig. 7. The MST result for pentagram network.

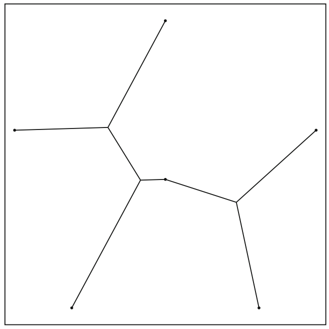

Fig. 8. ESMT by GeoSteiner for pentagram network.

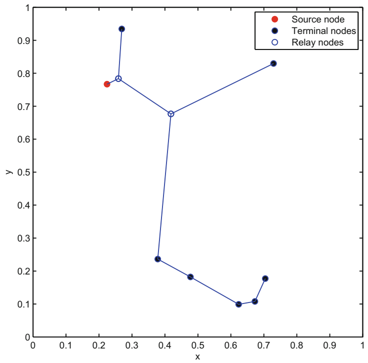

Fig. 9. CSIF result for random network when $N = 8$ and $M = 2$.

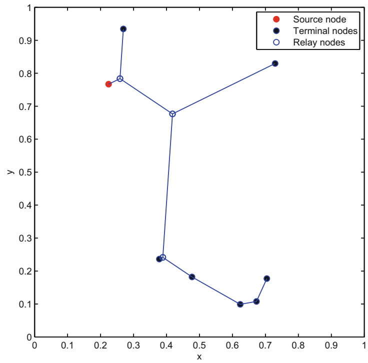

Fig. 10. SIF result for random network when $N = 8$.

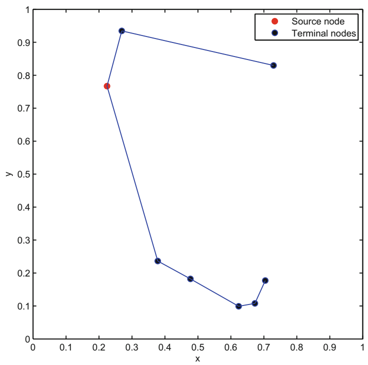

Fig. 11. MST result for random network when $N = 8$.

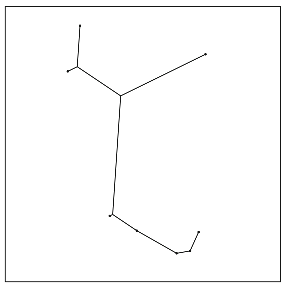

Fig. 12. The optimal ESMT by GeoSteiner for random network.

5 Conclusion

This work proposes a solution to the problem of Constrained Space Information Flow in multicast networks using a new $O(N^3 \log N)$ algorithm which takes into consideration a constraint on the number of relay nodes while computing the min-cost and the topology of the network for $N \geq 3$ terminal nodes in 2-D Euclidean space. The algorithm design is based on DT and LP techniques. The output of the algorithm is a min-cost multicast topology that consists of terminal (original) nodes and relay (additional) nodes. Our future work includes to apply Constrained Space Information Flow to wireless sensor networks.

Acknowledgments. This research was supported by National Natural Science Foundation of China (No. 61271227).

References

1. Ahlswede, R., Cai, N., Li, S.Y.R., Yeung, R.W.: Network information flow. IEEE Trans. Inf. Theory **46**(4), 1204–1216 (2000)
2. Li, Z., Wu, C.: Space information flow. Technical report, Department of Computer Science, University of Calgary (2011)
3. Li, Z., Wu, C.: Space information flow: multiple unicast. In: Proceedings of IEEE International Symposium on Information Theory (ISIT), pp. 1897–1901 (2012)
4. Gilbert, E.N., Pollak, H.O.: Steiner minimal trees. SIAM J. Appl. Math. **16**(1), 1–29 (1968)
5. Prim, R.C.: Shortest connection networks and some generalizations. Bell Syst. Tech. J. **36**(6), 1389–1401 (1957)
6. Huang, J., Yin, X., Zhang, X., Du, X., Li, Z.: On space information flow: single multicast. In: IEEE NetCod, pp. 1–6 (2013)
7. Maheshwar, S., Li, Z., Li, B.: Bounding the coding advantage of combination network coding in undirected networks. IEEE Trans. Inf. Theory **58**(2), 570–584 (2012)
8. Winter, P., Zachariasen, M.: Euclidean steiner minimum trees: an improved exact algorithm. Networks **30**(3), 149–166 (1997)
9. Huang, J., Li, Z.: A recursive partitioning algorithm for space information flow. In: IEEE GLOBECOM, pp. 1460–1465 (2014)
10. Yin, X., Wang, Y., Wang, X., Xue, X., Li, Z.: Min-cost multicast networks in euclidean space. In: IEEE ISIT, pp. 1316–1320 (2012)
11. Xiahou, T., Li, Z., Wu, C., Huang, J.: A geometric perspective to multiple-unicast network coding. IEEE Trans. Inf. Theory **60**(5), 2884–2895 (2014)
12. Hu, Y., Niu, D., Li, Z.: Internet video multicast via constrained space information flow. IEEE MMTC E-letter **9**(2), 17–19 (2014)
13. Uwitonze, A., Ye, Y., Huang, J., Cheng, W.: A heuristic algorithm on space information flow. In: 2015 International Conference on Computer Science and Applications, pp. 20–24 (2015)
14. Smith, J.M., Lee, D.T., Liebman, J.S.: An O (n log n) heuristic for steiner minimal tree problems on the euclidean metric. Networks **11**(1), 23–39 (1981)
15. Yue, M.: Minimum Network: The Steiner Tree Problem. Shanghai Scientific and Technical Publishers, Shanghai (2006)

16. Leach, G.: Improving worst-case optimal Delaunay triangulation algorithms. In: 4th Canadian Conference on Computational Geometry, pp. 340–346 (1992)
17. Beasley, J.E.: OR-Library: distributing test problems by electronic mail. J. Oper. Res. Soc. **41**(11), 1069–1072 (1990)
18. Waxman, B.M.: Routing of multipoint connections. IEEE J. Sel. Areas Commun. **6**(9), 1617–1622 (1988)