An Efficient Cost-Sharing Mechanism for the Prize-Collecting Steiner Forest Problem

A. Gupta * J. Könemann † S. Leonardi ‡ R. Ravi § G. Schäfer¶

Abstract

In an instance of the *prize-collecting Steiner forest* problem (PCSF) we are given an undirected graph $G = (V, E)$, non-negative edge-costs $c(e)$ for all $e \in E$, terminal pairs $R = \{(s_i, t_i)\}_{1 \le i \le k}$, and penalties $\pi_1, \ldots, \pi_k$. A feasible solution (F, Q) consists of a forest F and a subset Q of terminal pairs such that for all $(s_i, t_i) \in R$ either s_i, t_i are connected by F or $(s_i, t_i) \in Q$. The objective is to compute a feasible solution of minimum cost $c(F) + \pi(Q)$.

A game-theoretic version of the above problem has k players, one for each terminal-pair in R. Player i's ultimate goal is to connect s_i and t_i, and the player derives a privately held *utility* $u_i \ge 0$ from being connected. A service provider can connect the terminals s_i and t_i of player i in two ways: (1) by buying the edges of an s_i, t_i-path in G, or (2) by buying an alternate connection between s_i and t_i (maybe from some other provider) at a cost of π_i.

In this paper, we present a simple 3-budget-balanced and group-strategyproof mechanism for the above problem. We also show that our mechanism computes client sets whose social cost is at most $O(\log^2 k)$ times the minimum social cost of any player set. This matches a lower-bound that was recently given by Roughgarden and Sundararajan (STOC '06).

*School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA. Email: anupamg@andrew.cmu.edu.

†Department of Combinatorics and Optimization, University of Waterloo, Waterloo, ON N2L 3G1, Canada. Email: jochen@math.uwaterloo.ca. Research supported by NSERC grant no. 288340-2004 and by an IBM Faculty Award.

‡Dipartimento di Informatica e Sistemistica, University of Rome "La Sapienza", Via Salaria 113, 00198 Rome, Italy. Email: leon@dis.uniroma1.it. Part of this work was done while the author was visiting the School of Computer Science at Carnegie Mellon University.

§Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213, USA. Email: ravi@cmu.edu.

¶Institute of Mathematics, Technische Universität Berlin, Strasse des 17. Juni 136, 10623 Berlin, Germany. Email: schaefer@math.tu-berlin.de. Supported by the DFG Research Center Matheon "Mathematics for key technologies". Part of this work was done while the author was visiting the School of Computer Science at Carnegie Mellon University.

1 Introduction

In an instance of the *prize-collecting Steiner forest* problem (PCSF) we are given an undirected graph $G = (V, E)$ with edge costs $c : E \to \mathbb{R}^+$, a set of k terminal pairs $R = \{(s_i, t_i)\}_{1 \le i \le k}$, and penalties $\pi : R \to \mathbb{R}^+$. A feasible solution (F, Q) consists of a forest F and a subset Q of terminal pairs such that for all $(s_i, t_i) \in R$ either s_i, t_i are connected by F or $(s_i, t_i) \in Q$. The objective is to compute a feasible solution of minimum cost $c(F) + \pi(Q)$.

A game-theoretic version of the above problem has k players, one for each terminal-pair in R. We use U to denote the set of all players. Player i's ultimate goal is to connect s_i and t_i, and the player derives a privately held *utility* $u_i \ge 0$ from being connected. A service provider can connect the terminals s_i and t_i of player i in two ways: (1) by buying the edges of an s_i, t_i-path in G, or (2) by buying an alternate connection between s_i and t_i (maybe from some other provider) at a cost of π_i.

Formally, we are interested in finding a *cost-sharing mechanism* that first solicits bids $\{b_i\}_{i \in U}$ from all players. The mechanism then determines a set $S \subseteq U$ of players to service and computes a prize-collecting Steiner forest for the terminal set of these players. Finally, the mechanism needs to determine a payment $x_i(S) \le b_i$ for each of the players in S.

There are several desirable properties of a cost-sharing mechanism: a mechanism is called *strategyproof*, if bidding truthfully (i.e., announcing $b_i = u_i$) is a dominant strategy for all players. If this is true even if players are permitted to collude, then we call a mechanism *group-strategyproof*. A mechanism is *budget balanced* if the total cost $C(S)$ of servicing the players in S is at most the sum of the costs charged to the players in S, and it is *competitive* if the sum of all costs charged to the players in S does not exceed the cost of an optimal PCSF solution for S. A mechanism is called *efficient* if it selects a set S of players that maximizes $u(S) - C(S)$.

Classical results in economics [12, 24] state that budget balance and efficiency cannot be simultaneously achieved by any mechanism. Moreover, Feigenbaum et al. [10] recently showed that there is no strategyproof mechanism that always recovers a constant fraction of the maximum efficiency and a constant fraction of the incurred cost even for the simple fixed-tree multicast problem.

In light of these hardness results, most of the previous work on mechanism design concentrated on proper subsets of the above design goals. One notable class of such mechanisms are based on a framework due to Moulin and Shenker [22]. The authors showed that, given a budget balanced and *cross-monotonic cost sharing method* for the underlying problem, the well known Moulin mechanism [21] satisfies budget balance and group-strategyproofness. Moulin and Shenker's framework has recently been applied to game-theoretic variants of classical optimization problems such as fixed-tree multicast [2, 9, 10], submodular cost-sharing [22], Steiner trees [16, 17], facility location, single-source rent-or-buy network design [23, 20, 13] and Steiner forests [18]. Lower bounds on the budget balance factor that is achievable by a cross-monotonic cost sharing mechanism are given in [15, 19].

Very recently, Roughgarden and Sundararajan [25] introduced an alternative measure of efficiency that circumvents the intractability results in [10, 12, 24] at least partially. Let U be a universe of players and let C be a cost function on U that assigns to each subset $S \subseteq U$ a non-negative service cost $C(S)$. The authors define the *social cost* $\Pi(S)$ of a set $S \subseteq U$ as $\Pi(S) = u(U \setminus S) + C(S)$. A mechanism is said to be *α-approximate* if the set it outputs has social cost at most α times the minimum over all sets $S \subseteq U$. The intuition for this definition loosely comes from the fact that $u(U) - \Pi(S) = u(S) - C(S)$, which is the traditional definition of efficiency; since $u(U)$ is a constant, a set S has minimum social cost iff it has maximum efficiency.

Roughgarden and Sundararajan then developed a framework to quantify the extent to which a Moulin mechanism minimizes the social cost, and apply this framework to show that the Shapley mechanism is $O(\log k)$-approximate for submodular functions, and that the Steiner tree cost-shares of Jain and Vazirani [16] give a mechanism that is $O(\log^2 k)$-approximate.

1.1 Prize-Collecting Steiner Problems. Computing minimum-cost prize-collecting Steiner trees or forests is APX-complete [3, 5], and hence neither of the two problems admits a PTAS unless P = NP. The first constant-factor approximation for the prize-collecting Steiner tree problem was a LP-rounding based 3-approximation by Bienstock et al. [6], and this was improved to $2 - 1/k$ by Goemans and Williamson [11] using the primal-dual schema. One can easily modify the algorithm of Bienstock et al. to give a 3-approximation for the PCSF problem as well; in [14], Hajiaghayi and Jain refine Bienstock's LP rounding idea and obtain an LP-based 2.54 approximation for the problem. The authors also present a primal-dual combinatorial 3-approximation for the problem. This algorithm substantially deviates from the classical framework of Goemans and Williamson, requiring crucial use of Farkas' Lemma, wherein the dual variables are both increased and decreased along the execution of the algorithm.

1.2 Our Results and Techniques. The first contribution of this paper is the following:

THEOREM 1.1. *There is an efficiently computable cross-monotonic cost sharing method ξ^{GKLRS} for the prize-collecting Steiner forest problem that is 3-budget balanced.*

Our algorithm *GKLRS* is a natural extension of the primal-dual algorithm of Goemans and Williamson [11] for prize-collecting Steiner trees and the cross-monotonic cost sharing method *KLS* for Steiner forests presented in [18]. Despite its simplicity, our algorithm achieves the same approximation guarantee as [14].

Our second result bounds the social cost of the mechanism associated with the cost-sharing method.

THEOREM 1.2. *The Moulin mechanism $M(\xi^{GKLRS})$ driven by the cross-monotonic cost sharing method ξ^{GKLRS} is $\Theta(\log^2 k)$-approximate.*

This result is achieved in two steps. The first step is to show that if the Moulin mechanism $M(\xi^{KLS})$ is α-approximate then the mechanism $M(\xi^{GKLRS})$ given by our cross-monotonic cost-sharing method ξ^{GKLRS} is $3(1+\alpha)$-approximate for the prize-collecting Steiner forest game. As the second step, we show that the *KLS* mechanism is $O(\log^3 k)$-approximate for the Steiner Forest game. This is achieved by adding a novel methodological contribution to the framework proposed in [25]: we show that such a result can also be proved by embedding the graph distances into random HSTs [4, 8] rather than using the construction proposed by Roughgarden and Sundararajan. Independently, Chawla, Roughgarden and Sundararajan [7] have recently shown (using a more involved analysis) that *KLS* is $O(\log^2 k)$-approximate. We are optimistic that the general idea of reductions between cost-sharing mechanisms that we use in our proof will extend to the prize-collecting versions of other optimization problems.

1.3 Organization of the Paper. In Section 2 we introduce some notations used in the paper. In Section

3 we present the linear programming formulation for PCSF. Section 4 presents the cross-monotonic cost-sharing scheme *GKLRS* for PCSF. In Section 5 we prove the bound on the social cost for the *GKLRS* mechanism, whereas in Section 6 we prove the bound on the social cost for the Steiner forest mechanism *KLS*.

2 Preliminaries

Let U be a universe of players and let C be a cost function on U that assigns to each subset $S \subseteq U$ a non-negative cost $C(S)$. We assume that C is non-decreasing, i.e., for all $S \subseteq T$, $C(S) \leq C(T)$, and $C(\emptyset) = 0$.

2.1 Moulin Mechanisms. A *cost sharing method* ξ is an algorithm that, given any subset $S \subseteq U$ of players, computes a solution to service S and for each $i \in S$ determines a non-negative cost share $\xi_i(S)$. We say that ξ is *β-budget balance* if for every subset $S \subseteq U$,

$$\frac{1}{\beta} \cdot C(S) \leq \sum_{i \in S} \xi_i(S) \leq C(S).$$

A cost sharing method ξ is *cross-monotonic* if for any two sets S and T such that $S \subseteq T$ and any player $i \in S$ we have $\xi_i(S) \geq \xi_i(T)$.

Moulin and Shenker [22] showed that, given a budget balanced and cross-monotonic cost sharing method ξ for the underlying problem, the following cost sharing mechanism $M(\xi)$ satisfies budget-balance and group-strategyproofness: Initially, let $S = U$. If for each player $i \in S$ the cost share $\xi_i(S)$ is at most her bid b_i, we stop. Otherwise, remove from S all players whose cost shares are larger than their bids, and repeat. Eventually, let $\xi_i(S)$ be the costs that are charged to players in the final set S.

2.2 Approximating Social Cost. Roughgarden and Sundararajan [25] recently introduced an alternative notion of efficiency for cost sharing mechanisms: Every player $i \in U$ has a private utility u_i. For a set $S \subseteq U$, define $u(S) = \sum_{i \in S} u_i$. Define the *social cost* $\Pi(S)$ of a set $S \subseteq U$ as

$$\Pi(S) = u(U \setminus S) + C(S).$$

Definition 2.1. *Suppose S^M is the final set of players computed by the Moulin mechanism $M(\xi)$ on U. Then $M(\xi)$ is said to be* α-approximate *if*

$$\Pi(S^M) \leq \alpha \cdot \Pi(S) \quad \forall S \subseteq U.$$

Roughgarden and Sundararajan [25] proved that the Moulin mechanism $M(\xi)$ is $(\alpha + \beta)$-approximate and β-budget balanced if ξ is *α-summable* and β-budget balanced. The summability of a cost sharing method is defined as follows: Assume we are given an arbitrary permutation σ on the players in U and a subset $S \subseteq U$ of players. We assume that the players in S are ordered according to σ, i.e., $S = \{i_1, \ldots, i_{|S|}\}$ where $i_j \prec_\sigma i_k$ if and only if $1 \leq j < k \leq |S|$. We define $S_j \subseteq S$ as the (ordered) set of the first j players of S according to the order σ.

Definition 2.2. *A cost sharing method ξ is α-*summable *if for every ordering σ and every subset $S \subseteq U$*

$$\sum_{j=1}^{|S|} \xi_{i_j}(S_j) \leq \alpha \cdot C(S). \tag{2.1}$$

where S_j is the set of the first j players, and i_j is the j^{th} player according to the ordering σ.

3 LP Formulation

Subsequently, we slightly abuse notation by using R to refer to the set of terminal pairs and the set of terminals. For a terminal $u \in R$, let $\bar{u}$ be the mate of u, i.e., $(u, \bar{u}) \in R$. For a terminal pair $(u, \bar{u}) \in R$, define the *death time* as $\mathtt{d}(u, \bar{u}) = \frac{1}{2} d_G(u, \bar{u})$, where $d_G(u, \bar{u})$ is the cost of a shortest $u, \bar{u}$-path (with respect to c) in G.

Consider a cut $S \subseteq V$. We say S *separates* a terminal pair $(u, \bar{u}) \in R$ iff $|\{u, \bar{u}\} \cap S| = 1$. We also write $(u, \bar{u}) \odot S$ iff $(u, \bar{u})$ is separated by S. A cut S that separates at least one terminal pair is called a *Steiner cut*. Let $\mathcal{S}$ denote the set of all Steiner cuts. For a cut $S \subseteq V$, we use $\delta(S)$ to refer to the set of all edges $(u, v) \in E$ that *cross* S, i.e., $\delta(S) = \{(u, v) \in E : |\{u, v\} \cap S| = 1\}$.

A natural integer programming formulation for PCSF has a 0/1-variable x_e for all edges $e \in E$ and a 0/1-variable $x_{u\bar{u}}$ for all terminal pairs $(u, \bar{u}) \in R$. Variable $x_e = 1$ iff $e \in F$ and $x_{u\bar{u}} = 1$ iff $(u, \bar{u}) \in Q$. The following is an integer programming formulation for PCSF:

$$\text{(ILP)} \quad \min \sum_{e \in E} c(e) \cdot x_e + \sum_{(u,\bar{u}) \in R} \pi(u, \bar{u}) \cdot x_{u\bar{u}}$$

$$\text{s.t.} \sum_{e \in \delta(S)} x_e + x_{u\bar{u}} \geq 1 \quad \forall S \in \mathcal{S},\ \forall (u, \bar{u}) \odot S \tag{3.2}$$

$$x_e,\ x_{u\bar{u}} \in \{0, 1\} \quad \forall e \in E,\ \forall (u, \bar{u}) \in R.$$

We use OPT_R to refer to the cost of an optimal solution to this LP. Constraint (3.2) ensures that each Steiner cut $S \in \mathcal{S}$ is either crossed by an edge of F, or all separated terminal pairs $(u, \bar{u}) \odot S$ are part of Q.

The dual of the linear programming relaxation (LP) of (ILP) is as follows. We have a non-negative dual variable $\xi_{S,u\bar{u}}$ for all Steiner cuts $S \in \mathcal{S}$ and all pairs $(u,\bar{u}) \in R$ such that $(u,\bar{u}) \odot S$:

$$\text{(D)} \qquad \max \sum_{S\in\mathcal{S}} \sum_{(u,\bar{u})\odot S} \xi_{S,u\bar{u}}$$

$$\text{(3.3)} \qquad \text{s.t.} \sum_{S\in\mathcal{S}:e\in\delta(S)} \sum_{(u,\bar{u})\odot S} \xi_{S,u\bar{u}} \le c(e) \quad \forall e \in E$$

$$\text{(3.4)} \qquad \sum_{S\in\mathcal{S}:S\odot(u,\bar{u})} \xi_{S,u\bar{u}} \le \pi(u,\bar{u}) \quad \forall (u,\bar{u}) \in R$$

$$\xi_{S,u\bar{u}} \ge 0 \quad \forall S \in \mathcal{S},\ (u,\bar{u}) \odot S.$$

It is convenient to associate a dual solution $\{\xi_{S,u\bar{u}}\}_{S\in\mathcal{S},(u,\bar{u})\odot S}$ with a set of dual values $\{y_S\}_{S\in\mathcal{S}}$ for all Steiner cuts $S \in \mathcal{S}$. To this aim, we define the dual y_S of a Steiner cut $S \in \mathcal{S}$ simply as the total cost share of all its separated terminal pairs:

$$y_S = \sum_{(u,\bar{u})\odot S} \xi_{S,u\bar{u}}.$$

We can think of $\xi_{S,u\bar{u}}$, $(u,\bar{u}) \odot S$, as a *cost share* that terminal pair $(u,\bar{u})$ receives from dual y_S of S. Define the total cost share of $(u,\bar{u})$ as

$$\xi_{u\bar{u}} = \sum_{S\in\mathcal{S}:S\odot(u,\bar{u})} \xi_{S,u\bar{u}}.$$

Clearly, with these definitions

$$\sum_{S\in\mathcal{S}} y_S = \sum_{(u,\bar{u})\in R} \xi_{u\bar{u}}.$$

Constraint (3.3) of LP (D) requires that for every edge $e \in E$, the total dual of all Steiner cuts $S \in \mathcal{S}$ that cross e is at most the cost $c(e)$ of this edge. Constraint (3.4) states that the total cost share $\xi_{u\bar{u}}$ of terminal pair $(u,\bar{u})$ is at most its penalty $\pi(u,\bar{u})$.

4 A Cross-Monotonic Algorithm for the PCSF Problem

Our algorithm *GKLRS* for the prize-collecting Steiner forest problem is a primal-dual algorithm, that is, it maintains a primal solution $\{x_e, x_{u\bar{u}}\}_{e\in E,(u,\bar{u})\in R}$ together with a set of dual values $\{y_S\}_{S\in\mathcal{U}}$ (the definition of the set $\mathcal{U}$ is given below). The primal solution is a 0/1-solution that is infeasible for (LP) initially. Throughout the execution of *GKLRS*, the degree of infeasibility of this solution is decreased successively until eventually, we obtain a feasible solution for (LP).

A subtle point of our algorithm is that it does not produce a set of dual values $\{y_S\}_{S\in\mathcal{U}}$ that corresponds to a feasible solution for (D). There are two reasons for this. First, we also raise dual values y_S of cuts S that do not correspond to Steiner cuts. We use $\mathcal{U}$ to refer to the set of all cuts that are raised throughout the execution of *GKLRS*. As a consequence, a terminal pair $(u,\bar{u})$ may receive cost share $\xi_{S,u\bar{u}}$ from a non-Steiner cut $S \in \mathcal{U}\backslash\mathcal{S}$. Second, a terminal pair $(u,\bar{u})$ may also receive cost share $\xi_{S,u\bar{u}}$ from a cut S that does not separate $(u,\bar{u})$. However, *GKLRS* maintains the invariant that a terminal pair $(u,\bar{u})$ only receives cost share from cuts $S \in \mathcal{U}$ that either separate or entirely contain $(u,\bar{u})$, i.e., $(u,\bar{u}) \odot S$ or $\{u,\bar{u}\} \subseteq S$.

We can view the execution of *GKLRS* as a process over time. Initially, at time $\tau = 0$, $x_e^\tau = 0$ for all $e \in E$, $x_{u\bar{u}}^\tau = 0$ for all $(u,\bar{u}) \in R$ and $y_S^\tau = 0$ for all $S \in \mathcal{U}$. Let F^τ be the forest that corresponds to $\{x_e^\tau\}_{e\in E}$, i.e., $F^\tau = \{e \in E : x_e^\tau = 1\}$. Similarly, let Q^τ be the set of all terminal pairs $(u,\bar{u}) \in R$ such that $x_{u\bar{u}}^\tau = 1$.

We define $\bar{F}^\tau$ as the set of all edges that are tight at time τ, i.e.,

$$\bar{F}^\tau = \{e \in E : \sum_{S\in\mathcal{U}} y_S^\tau = c(e)\}.$$

We use the term *moat* to refer to a connected component M^τ in $\bar{F}^\tau$. A moat M^τ defines a cut S which is simply the set of vertices spanned by M^τ. At time τ, we increase the duals of all cuts defined by moats $M^\tau \in \bar{F}^\tau$ that are *active* at time τ. The notion of activity will be defined shortly. These duals are increased simultaneously and by the same amount. Subsequently, we also say that we *grow* all active moats in $\bar{F}^\tau$ at time τ. Moreover, it is convenient to regard the growing of moats as being identical to increasing the duals.

4.1 Activity Notion. We call a terminal pair $(u,\bar{u}) \in R$ *active* at time τ if

$$\text{(4.5)} \qquad \xi_{u\bar{u}}^\tau < \pi(u,\bar{u}) \quad \text{and} \quad \tau < \mathtt{d}(u,\bar{u}).$$

If the above conditions do not hold, we say that $(u,\bar{u})$ is *inactive* at time τ. Let $\tau_{u\bar{u}}$ be the first time when $(u,\bar{u})$ becomes inactive. Observe that by definition (4.5), a terminal pair $(u,\bar{u})$ remains inactive at all times $\tau > \tau_{u\bar{u}}$. A terminal $u \in R$ is active at time τ if its pair $(u,\bar{u})$ is active at this time. Let $\mathcal{A}^\tau$ be the set of all terminals that are active at time τ.

We say that a moat $M^\tau \in \bar{F}^\tau$ is active at time τ if it contains at least one active terminal, i.e., $M^\tau \cap \mathcal{A}^\tau \neq \emptyset$. The growth of an active moat M^τ is shared evenly among all active terminals in M^τ. Let $M^\tau(u)$ denote the moat in $\bar{F}^\tau$ that contains terminal $u \in R$. More formally, we define the cost share $\xi_u^{\tau'}$ of a terminal $u \in R$ at time $\tau' \le \tau_{u\bar{u}}$ as follows:

$$\text{(4.6)} \qquad \xi_u^{\tau'} = \int_0^{\tau'} \frac{1}{|M^\tau(u) \cap \mathcal{A}^\tau|} d\tau.$$

Let $\xi_u^{\tau'} = \xi_u^{\tau_{u\bar{u}}}$ for all $\tau' > \tau_{u\bar{u}}$. Moreover, we define $\xi_{u\bar{u}}^\tau = \xi_u^\tau + \xi_{\bar{u}}^\tau$.

Observe that the total contribution to the cost share of a terminal pair $(u, \bar{u})$ within ϵ time units is at most 2ϵ. Also, note that $(u, \bar{u})$ may receive cost share from a moat M^τ that contains u and $\bar{u}$.

The following fact follows immediately from definitions (4.5) and (4.6).

FACT 4.1. *For all terminal pairs $(u, \bar{u}) \in R$, $\xi_{u\bar{u}} \le \min\{\pi(u, \bar{u}), 2\mathtt{d}(u, \bar{u})\}$.*

Since at any point of time, the growth of all active moats is shared among active terminals, the following must hold true.

FACT 4.2. *For every time $\tau \ge 0$,*

$$\sum_{S \in \mathcal{U}} y_S^\tau = \sum_{(u,\bar{u}) \in R} \xi_{u\bar{u}}^\tau.$$

We say that two active moats M_1 and M_2 *collide* at time τ if their vertices are contained in the same connected component of $\bar{F}^{\tau'}$ iff $\tau' \ge \tau$. In this case, we add a cheapest collection of edges to F^τ s.t. all active vertices of M_1 and M_2 are in the same connected component of $F^{\tau'}$ for all $\tau' \ge \tau$.

Suppose a terminal pair $(u, \bar{u}) \in R$ becomes inactive at time $\tau = \tau_{u\bar{u}}$ because it reaches its penalty, i.e., $\xi_{u\bar{u}}^\tau = \pi(u, \bar{u})$. We then add $(u, \bar{u})$ to Q^τ. Since $(u, \bar{u})$ remains inactive after time $\tau_{u\bar{u}}$, the following fact holds true.

FACT 4.3. *Let Q be the final set of terminal pairs computed by GKLRS. Then*

$$\sum_{(u,\bar{u}) \in Q} \pi(u, \bar{u}) = \sum_{(u,\bar{u}) \in Q} \xi_{u\bar{u}}$$

Suppose a terminal pair $(u, \bar{u})$ becomes inactive at time $\mathtt{d}(u, \bar{u})$. The next fact shows that $(u, \bar{u})$ must then be connected in F.

FACT 4.4. *Let terminal pair $(u, \bar{u})$ become inactive just after time $\mathtt{d}(u, \bar{u})$. Then u and $\bar{u}$ are connected in F.*

Proof. Let $P_{u\bar{u}}$ be a shortest $u, \bar{u}$-path in G. Path $P_{u\bar{u}}$ becomes tight at time $\tau \le \mathtt{d}(u, \bar{u})$ and both u and $\bar{u}$ are active at this time. Thus either u and $\bar{u}$ are already connected in F^τ or $P_{u\bar{u}}$ is added to F^τ.

Observe that the last fact also establishes correctness of *GKLRS*: The final solution (F, Q) computed by *GKLRS* is a feasible solution for the given prize-collecting Steiner forest instance.

Subsequently, we use $\xi^{GKLRS}(S)$ to refer to final cost shares computed by *GKLRS* when run on terminal set $S \subseteq R$. We also identify the player set U with the terminal-pair set R.

4.2 Cross-Monotonicity. We compare the execution of *GKLRS* on terminal set R with the one on terminal set $R_{-st} = R \setminus \{(s, t)\}$ for any $(s, t) \in R$. We use $\mathcal{G}_{-st}$ ($\mathcal{G}$ = *GKLRS*, F, $\bar{F}$, M, etc.) to refer to $\mathcal{G}$ in the run of *GKLRS* on R_{-st}. For notational convenience, let $\xi_{-st}(u, \bar{u})$ refer to the cost share of $(u, \bar{u})$ in the run of *GKLRS* on R_{-st} and let $\xi(u, \bar{u})$ refer to the respective cost share in *GKLRS* on R.

LEMMA 4.1. *Consider the execution of GKLRS on R and R_{-st}, respectively. The following holds for every time $\tau \ge 0$:*

1. *$\bar{F}_{-st}^\tau$ is a refinement of $\bar{F}^\tau$, i.e., $\bar{F}_{-st}^\tau \subseteq \bar{F}^\tau$.*
2. *For all $(u, \bar{u}) \in R_{-st}$, $\xi_{-st}^\tau(u, \bar{u}) \ge \xi^\tau(u, \bar{u})$.*

Proof. We prove the lemma by induction over time τ. Clearly, the lemma holds at time $\tau = 0$. Suppose the lemma holds at time τ.

The only moats that may potentially violate the claim $\bar{F}_{-st}^{\tau+\epsilon} \subseteq \bar{F}^{\tau+\epsilon}$ at time $\tau + \epsilon$ for some $\epsilon > 0$, are those that are active at time τ in *GKLRS$_{-st}$*. Let $M_{-st} \in \bar{F}_{-st}^\tau$ be a moat that is active at time τ. By the induction hypothesis, there exists a moat $M \in \bar{F}^\tau$ such that $M_{-st} \subseteq M$. We argue that M must be active at time τ in *GKLRS*.

Since M_{-st} is active at time τ, there must exist a terminal $u \in M_{-st}$ such that $\pi(u, \bar{u}) - \xi_{-st}^\tau(u, \bar{u}) > 0$ and $\tau < \mathtt{d}(u, \bar{u})$. By our induction hypothesis,

$$\pi(u, \bar{u}) - \xi^\tau(u, \bar{u}) \ge \pi(u, \bar{u}) - \xi_{-st}^\tau(u, \bar{u}) > 0.$$

Therefore, M must be active at time τ too. This proves the first part of the lemma.

It remains to be shown that $\xi_{-st}^{\tau+\epsilon}(u, \bar{u}) \ge \xi^{\tau+\epsilon}(u, \bar{u})$ for all $(u, \bar{u}) \in R_{-st}$. Observe that all terminal pairs that are inactive at time τ do not receive any further cost share. Consider a terminal pair $(u, \bar{u}) \in R_{-st}$ that is active at time τ in *GKLRS$_{-st}$* and let $M_{-st}^\tau(u)$ be the moat of u at time τ. From the discussion above, we know that every terminal pair $(v, \bar{v}) \in R_{-st}$ that is active at time τ in *GKLRS$_{-st}$* must be active at time τ in *GKLRS*, i.e., $\mathcal{A}_{-st}^\tau \subseteq \mathcal{A}^\tau$. By our induction hypothesis, moat $M_{-st}^\tau(u)$ is contained in the moat $M^\tau(u) \in \bar{F}^\tau$ of u in *GKLRS*. Therefore, $|M_{-st}^\tau(u) \cap \mathcal{A}_{-st}^\tau| \le |M^\tau(u) \cap \mathcal{A}^\tau|$. Thus, the additional cost share that $(u, \bar{u})$ receives in the time interval $(\tau, \tau + \epsilon]$ in *GKLRS$_{-st}$* is at least as large as the one it receives in *GKLRS*.

4.3 Competitiveness. We next show that the total cost share of all terminal pairs is at most the cost of an optimal solution to the prize-collecting Steiner forest instance. The following proof is similar to the one presented in [18].

LEMMA 4.2. *Let (F^*, Q^*) be an optimal solution to the prize-collecting Steiner forest instance with terminal pair set R. The cost shares ξ computed by* GKLRS *for R satisfy*

$$\sum_{(u,\bar{u})\in R} \xi_{u\bar{u}} \le c(F^*) + \pi(Q^*).$$

Proof. Consider a separated terminal pair $(u, \bar{u}) \in Q^*$. By Fact 4.1, we have

$$\sum_{(u,\bar{u})\in Q^*} \xi_{u\bar{u}} \le \pi(Q^*).$$

It remains to be shown that the total cost share of all terminal pairs $(u, \bar{u}) \in R \setminus Q^*$ is bounded by $c(F^*)$.

Consider a connected component $T \in F^*$ and let $R(T)$ be the set of terminal pairs that are connected by T. We prove that

$$\sum_{(u,\bar{u})\in R(T)} \xi_{u\bar{u}} \le c(T). \tag{4.7}$$

The lemma follows by summing over all connected components $T \in F^*$.

We define $\mathcal{M}^\tau(T) \subseteq \bar{F}^\tau$ as the set of moats at time τ that contain at least one active terminal of $R(T)$, i.e.,

$$\mathcal{M}^\tau(T) = \{M^\tau(u) \,:\, u \in R(T) \cap \mathcal{A}^\tau\}.$$

Among all terminal pairs in $R(T)$, let $(w, \bar{w})$ be a pair that is active longest. By our definition of activity in (4.5), all terminal pairs in $R(T)$ are inactive after time $\mathtt{d}(w, \bar{w})$. We show that the total growth of $\mathcal{M}^\tau(T)$ for all $\tau \in [0, \mathtt{d}(w, \bar{w})]$ is at most $c(T)$. This implies (4.7).

At any time τ, the moats in $\mathcal{M}^\tau(T)$ are disjoint. Moreover, T connects all terminals in $R(T)$. Thus, if there exists a moat $M^\tau \in \mathcal{M}^\tau(T)$ that intersects an edge of T then each moat in $\mathcal{M}^\tau(T)$ must intersect an edge of T; we say that the moats in $\mathcal{M}^\tau(T)$ *load* T. Moreover, each moat M^τ loads a different part of T. Thus, the total growth of moats in $\mathcal{M}^\tau(T)$ for all τ at which $\mathcal{M}^\tau(T)$ loads T is at most $c(T)$.

Let $\tau_0 \le \mathtt{d}(w, \bar{w})$ be the first time such that $\mathcal{M}^{\tau_0}(T)$ does not load T. If $\mathcal{M}^{\tau_0}(T) = \emptyset$, we are done. Otherwise, we must have that $\mathcal{M}^{\tau_0}(T) = \{M^{\tau_0}\}$ and $T \subseteq M^{\tau_0}$. The additional growth of M^τ for all times $\tau \in [\tau_0, \mathtt{d}(w, \bar{w})]$ is at most $\mathtt{d}(w, \bar{w}) - \tau_0$. Since w and $\bar{w}$ are connected by T, this additional growth is at most $\mathtt{d}(w, \bar{w}) \le c(T)/2$. This gives an upper bound of $\frac{3}{2}c(T)$ on the total cost shares of pairs in $R(T)$.

The following refined argument proves (4.7). Let $P_{w\bar{w}}$ be the unique $w, \bar{w}$-path in T. Define $\mathcal{M}^\tau_{w\bar{w}} \subseteq \mathcal{M}^\tau(T)$ as the set of active moats different from $M^\tau(w)$ and $M^\tau(\bar{w})$ that load $P_{w\bar{w}}$ at time $\tau < \tau_0$, i.e.,

$$\mathcal{M}^\tau_{w\bar{w}} = \{M^\tau \in \mathcal{M}^\tau(T) \setminus \{M^\tau(w), M^\tau(\bar{w})\} \,:\, \delta(M^\tau) \cap P_{w\bar{w}} \ne \emptyset\}.$$

Define the *degree* $\mathtt{dg}(M^\tau)$ of a moat $M^\tau \in \mathcal{M}^\tau_{w\bar{w}}$ as

$$\mathtt{dg}(M^\tau) = |\delta(M^\tau) \cap P_{w\bar{w}}|.$$

PROPOSITION 4.1. *Consider a time $\tau < \tau_0$ and a moat $M^\tau \in \mathcal{M}^\tau_{w\bar{w}}$. Then $\mathtt{dg}(M^\tau) \ge 2$.*

Proof. Both $M^\tau(w)$ and $M^\tau(\bar{w})$ are active at time $\tau < \tau_0$ and thus $\{M^\tau(w), M^\tau(\bar{w})\} \subseteq \mathcal{M}^\tau(T)$ (possibly $M^\tau(w) = M^\tau(\bar{w})$). By definition of $\mathcal{M}^\tau_{w\bar{w}}$, $M^\tau \in \mathcal{M}^\tau(T)$ and $M^\tau \notin \{M^\tau(w), M^\tau(\bar{w})\}$. Furthermore, M^τ is disjoint from all other moats in $\mathcal{M}^\tau(T)$. Suppose $|M^\tau \cap P_{w\bar{w}}| = 1$. But then, moat M^τ must contain w or $\bar{w}$. This contradicts the disjointness of M^τ and $\{M^\tau(w), M^\tau(\bar{w})\}$.

By our choice of $(w, \bar{w}) \in R(T)$ as the terminal pair with largest activity time and by our assumption that $\mathcal{M}^{\tau_0}(T) \ne \emptyset$ it follows that both, $M^\tau(w)$ and $M^\tau(\bar{w})$ are active for all $0 \le \tau \le \tau_0$. We define $l_{w\bar{w}}$ as the total dual growth of the moats containing w and $\bar{w}$ up to time τ_0. Formally, let

$$\delta^\tau_{w\bar{w}} = \begin{cases} 2 & : \quad M^\tau(w) \ne M^\tau(\bar{w}) \\ 1 & : \quad \text{otherwise} \end{cases}$$

and

$$l_{w\bar{w}} = \int_0^{\tau_0} \delta^\tau_{w\bar{w}} d\tau.$$

It follows that the cost of path $P_{w\bar{w}}$ is at least

$$l_{w\bar{w}} + \int_0^{\tau_0} \sum_{M^\tau \in \mathcal{M}^\tau_{w\bar{w}}} \mathtt{dg}(M^\tau) d\tau.$$

We let $\mathtt{sl}_{w\bar{w}}$ be the difference between $c(P_{w\bar{w}})$ and the above term and obtain

$$c(P_{w\bar{w}}) = l_{w\bar{w}} + \mathtt{sl}_{w\bar{w}} + \int_0^{\tau_0} \sum_{M^\tau \in \mathcal{M}^\tau_{w\bar{w}}} \mathtt{dg}(M^\tau) d\tau. \tag{4.8}$$

We define the total growth $y^{\tau_0}(T)$ produced by terminal pairs in $R(T)$ until time τ_0 as follows:

$$y^{\tau_0}(T) = \int_0^{\tau_0} |\mathcal{M}^\tau(T)| d\tau.$$

At all times $\tau \le \tau_0$, each moat in $\mathcal{M}^\tau(T)$ loads at least one distinct edge of T; those in $\mathcal{M}^\tau_{w\bar{w}}$ load at least two edges of T. Thus, we have

$$c(T) \ge y^{\tau_0}(T) + \mathtt{sl}_{w\bar{w}} + \int_0^{\tau_0} \sum_{M^\tau \in \mathcal{M}^\tau_{w\bar{w}}} (\mathtt{dg}(M^\tau) - 1) d\tau. \tag{4.9}$$

The additional growth between time τ_0 and $\mathtt{d}(w,\bar{w})$ is at most $\mathtt{d}(w,\bar{w}) - \tau_0$. Using (4.8), we obtain

$$\begin{aligned}\mathtt{d}(w,\bar{w}) - \tau_0 \le \frac{l_{w\bar{w}}}{2} - \tau_0 + \frac{\mathtt{sl}_{w\bar{w}}}{2} \\ + \int_0^{\tau_0} \sum_{M^\tau \in \mathcal{M}^\tau_{w\bar{w}}} \frac{\mathtt{dg}(M^\tau)}{2} d\tau \\ \le \frac{\mathtt{sl}_{w\bar{w}}}{2} + \int_0^{\tau_0} \sum_{M^\tau \in \mathcal{M}^\tau_{w\bar{w}}} (\mathtt{dg}(M^\tau) - 1) d\tau,\end{aligned}$$

where we exploit that $\mathtt{dg}(M^\tau) \ge 2$ for all $M^\tau \in \mathcal{M}^\tau_{w\bar{w}}$ and the fact that $l_{w\bar{w}} \le 2\tau_0$. The last inequality together with (4.9) proves that the total growth is at most $c(T)$.

4.4 Cost Recovery

LEMMA 4.3. *Let (F,Q) be the solution and ξ be the cost shares computed by GKLRS on terminal pair set R, respectively. Then*

$$c(F) + \pi(Q) \le 3 \sum_{(u,\bar{u}) \in R} \xi_{u\bar{u}}.$$

Proof. Following the proof of Agrawal, Klein and Ravi [1], the cost of the constructed forest F satisfies

$$c(F) \le 2 \sum_{(u,\bar{u}) \in R} \xi_{u\bar{u}}.$$

Moreover, by Fact 4.3

$$\pi(Q) = \sum_{(u,\bar{u}) \in Q} \xi_{u\bar{u}}$$

and hence $c(F) + \pi(Q) \le 3 \sum_{(u,\bar{u}) \in R} \xi_{u\bar{u}}$.

5 Efficiency of *GKLRS*

In a very recent work, Chawla et al. [7] showed that the cost shares computed by *KLS* are also $O(\log^2 k)$-approximate. (A simple proof that they are $O(\log^3 k)$-approximate is given in Section 6.) In this paper, we extend this result to the prize-collecting Steiner forest (PCSF) game. We show that the approximability of *GKLRS* can be reduced to the one of *KLS*.

THEOREM 5.1. *If the mechanism $M(\xi^{KLS})$ is α-approximate then the mechanism $M(\xi^{GKLRS})$ is $3(1+\alpha)$-approximate.*

We will prove this theorem in the rest of this section. The following fact will be useful, and is easily proved.

FACT 5.1. *Given a cross-monotonic cost-sharing method ξ, the final set of players output by the Moulin mechanism $M(\xi)$ is independent of the order of eviction.*

The following lemma will allow us to partition the universe of players into two groups and to argue about each of them separately; due to space restrictions, we omit its proof.

LEMMA 5.1. *Consider a universe U of players, along with a non-decreasing cost function C and a β-budget balanced and cross-monotonic cost-sharing method ξ. Given a partition of U into two parts U_1 and U_2, if the Moulin mechanism on sub-universe U_i is α_i-approximate for all $i \in \{1,2\}$ with respect to the induced cost-sharing method $\xi|_{U_i}$ and the cost function $C|_{U_i}$, then the Moulin mechanism is $(\alpha_1+\alpha_2)\beta$-approximate for the entire set U with respect to ξ and C.*

Armed with the above lemma, let us consider the universe of players U for the *GKLRS* instance, and divide them into two parts thus:

- The "high-utility" set U_1 are those players $i \in U$ with utility $u_i \ge \pi_i$.
- The "low-utility" set U_2 are the remaining players $i \in U$ with $u_i < \pi_i$.

We now show that ξ^{GKLRS} on the sub-universes U_1 and U_2 is 1-approximate and α-approximate, respectively. This together with Lemma 5.1 and the fact that *GKLRS* is 3-budget balanced (Lemma 4.3) proves that *GKLRS* is $3(1+\alpha)$-approximate.

We first prove the following High-Utility-Lemma:

LEMMA 5.2. *The mechanism $M(\xi^{GKLRS})$ is 1-approximate when restricted to the players in the high-utility set U_1.*

Proof. By Fact 4.1, $\xi_i^{GKLRS}(S) \le \pi_i$ for every set $S \subseteq U$ and every $i \in S$. Since $u_i \ge \pi_i \ge \xi_i^{GKLRS}(S)$ for any $S \subseteq U_1$ and $i \in S$, the players in U_1 will never be rejected by the mechanism $M(\xi^{GKLRS})$ when run on U_1. Moreover, the set achieving the optimal social cost is also U_1, and hence the Moulin mechanism gives the social optimum on the high-utility set.

We show that for low-utility players $S \subseteq U_2$ the two runs of *GKLRS*(S) and *KLS*(S) are identical up to a certain point of time.

LEMMA 5.3. *Let $S \subseteq U_2$. Define τ_0 as the first point of time τ at which $\xi_i^{\tau,GKLRS}(S) = \pi_i$ for some player $i \in S$; let $\tau_0 = \infty$ if no such time exists. Then for all $\tau \in [0,\tau_0)$ and every player $j \in S$ it holds that j is active at time τ in GKLRS(S) iff j is active at time τ in KLS(S); in particular, this implies*

$$\xi_j^{\tau,GKLRS}(S) = \xi_j^{\tau,KLS}(S) \quad \forall \tau \in [0,\tau_0),\ \forall j \in S.$$

Proof. A necessary condition for j being active at time τ in GKLRS(S) is that $\tau \leq \mathtt{d}(s_j, t_j)$. Thus, j is active at time τ in KLS(S) if j is active at this time in GKLRS(S). Next, suppose j is active at time τ in KLS(S) and thus $\tau \leq \mathtt{d}(s_j, t_j)$. Since $\tau < \tau_0$, we have $\xi_i^{\tau,GKLRS}(S) < \pi_i$ for all $i \in S$; in particular this also holds for player j. Thus, j is active at time τ in GKLRS(S).

Suppose we compare the runs of the Moulin mechanism corresponding to the two different cost-sharing mechanisms ξ^{GKLRS} and ξ^{KLS} with the same set of low-utility players $S \subseteq U_2$. An immediate consequence of Lemma 5.3 is that as long as some player is eliminated in either of the runs of the Moulin mechanisms, there must be a player that the mechanisms could eliminate in *both the runs.*

COROLLARY 5.1. *Fix some $S \subseteq U_2$. Suppose there is a player $j \in S$ with $\xi_j^{GKLRS}(S) > u_j$ or $\xi_j^{KLS}(S) > u_j$. Then there is a player i such that $\xi_i^{GKLRS}(S) > u_i$ and $\xi_i^{KLS}(S) > u_i$.*

Proof. Let τ_0 be as defined in Lemma 5.3. The claim clearly holds if $\tau_0 = \infty$ as all cost shares in GKLRS(S) and KLS(S) are the same. Otherwise, there exists some player $i \in S$ and some $\tau_0 = \tau$ such that $\xi_i^{\tau,GKLRS}(S) = \pi_i$. Lemma 5.3 then implies that $\xi_i^{\tau,GKLRS}(S) = \xi_i^{\tau,KLS}(S) = \pi_i > u_i$.

The next lemma essentially shows that the prizes π_i play no role for the low-utility players U_2.

LEMMA 5.4. *When starting with a set of low-utility players U_2, the final output $S^{M,GKLRS} \subseteq U_2$ of the Moulin mechanism $M(\xi^{GKLRS})$ is identical to the output $S^{M,KLS} \subseteq U_2$ of the Moulin mechanism $M(\xi^{KLS})$.*

Proof. Corollary 5.1 states that we can always identify a player $i \in S$ that we may evict in both runs of $M(\xi^{GKLRS})$ and $M(\xi^{KLS})$ as long as some player is eliminated in either of the runs of the Moulin mechanism. We can then eliminate player i in both the runs and use induction to show that both runs end with the same players if we make the right choices. However, Fact 5.1 implies that *any* choices would lead to the same outputs, as we claim.

We can now prove the following Low-Utility Lemma:

LEMMA 5.5. *Restricting our attention to the low-utility set U_2, the mechanism $M(\xi^{GKLRS})$ is α-approximate if the mechanism $M(\xi^{KLS})$ is α-approximate.*

Proof. On the low-utility players, the solution with the optimal social cost for PCSF would never service a player i by paying her penalty π_i, since it would be better to reject the player and pay $u_i < \pi_i$. This implies that the optimal social cost Π^*_{PCSF} for PCSF and and the optimal social cost Π^*_{SF} for SF are the same on U_2. Also note that for every player set S the cost $OPT_{PCSF}(S)$ of an optimal PCSF solution for S is at most the cost $OPT_{SF}(S)$ of an optimal SF solution. Let Π_{PCSF} and Π_{SF} denote the social cost with respect to PCSF and SF, respectively. Given these facts together with the fact that $M(\xi^{GKLRS})$ and $M(\xi^{KLS})$ output the same set S^M on the low-utility instances, we conclude that

$$\begin{aligned}\Pi_{PCSF}(S^M) &= u(U_2 \setminus S^M) + OPT_{PCSF}(S^M)\\ &\leq u(U_2 \setminus S^M) + OPT_{SF}(S^M)\\ &= \Pi_{SF}(S^M) \leq \alpha \cdot \Pi^*_{SF} = \alpha \cdot \Pi^*_{PCSF}.\end{aligned}$$

6 Efficiency of *KLS*

As mentioned earlier, Chawla et al. [7] recently proved that the cost shares of KLS are $O(\log^2 k)$-approximate. Here we give a weaker result (but with a simple proof).

THEOREM 6.1. *The cost shares ξ^{KLS} computed by KLS are $O(\log^3 k)$-summable.*

Due to space restrictions, we sketch the ideas here only; details are given in the full version of the paper. Subsequently, we drop the superscript KLS. Recall that for every ordering σ and every subset $S \subseteq U$ of terminal pairs, we need to prove that

$$\sum_{i=1}^{|S|} \xi_i(S_i) = O(\log^3 k \cdot OPT(S)), \tag{6.10}$$

where $OPT(S)$ is the minimum Steiner forest cost for terminal set S. As before, S_i is the set of the first i terminal pairs in S and $\xi_i(S_i)$ refers to the cost share of (s_i, t_i) computed by KLS when run on terminal pair set S_i.

We partition terminal pairs according to their death times into *classes*. A terminal pair $(s_i, t_i) \in S$ is of *class* $r \geq 0$ iff $\mathtt{d}(s_i, t_i) \in (2^{r-1}, 2^r]$. Suppose there are at most $O(\log k)$ non-empty classes; we show in the full version of the paper how to circumvent this assumption. Exploiting the cross-monotonicity of ξ, one can easily verify that ξ is $O(\log^3 k)$-summable for S if ξ is $O(\log^2 k)$-summable for each class. The following Rounding Lemma states that we may even assume that the death times of all terminal pairs are equal.

LEMMA 6.1. *Suppose ξ^{KLS} is α-summable if all death times are equal to 2^r for some $r \geq 0$. Then ξ^{KLS}*

is $O(\alpha)$-summable if all death times are in the range $(2^{r-1}, 2^r]$.

Subsequently, we assume that every terminal pair in S has death time $D = 2^r$. Consider an optimal Steiner forest F^* for S. The forest F^* naturally induces a partition of S. Let $S(T)$ be the set of terminal pairs that are connected by tree $T \in F^*$. For a terminal pair $(s_i, t_i) \in S$ that is part of tree $T \in F^*$, define $S_i(T)$ as the set of terminal pairs that are also contained in T and precede (s_i, t_i), i.e., $S_i(T) = S_i \cap S(T)$. We show for each tree $T \in F^*$ separately that

$$\sum_{(s_i,t_i)\in S(T)} \xi_i(S_i(T)) = O(\log^2 k \cdot c(T)). \tag{6.11}$$

Summing over all trees and exploiting cross-monotonicity of ξ, then shows $O(\log^2 k)$-summability of ξ on S.

Fix a tree $T \in F^*$. We construct a rooted tree $T' = (V', E')$ and a non-negative length function $\ell : E' \to \mathbb{R}^+$ on the edges of T' satisfying the following properties:

1. The leaves of T' are the terminals in $S(T)$.
2. For every two terminals that are contained in the subtree $T'(e)$ for some $e \in E'$, their distance in G is at most $\ell(e)$, i.e., $d_G(u, v) \leq \ell(e)$ for all $u, v \in S(T) \cap T'(e)$.
3. For every path $P_{ur} = (e_1, \ldots, e_m)$ from terminal $u \in S(T)$ to the root r of T', we have
   (a) $\ell(e_1) = 1$,
   (b) $\ell(e_j) = 2\ell(e_{j-1})$ for all $1 < j \leq m$, and
   (c) $\ell(e_m) \geq D$.
4. The total length of T' is at most $O(\log |S(T)|)$ times the total cost of T, i.e., $\ell(T') = O(\log(|S(T)|) \cdot c(T))$.

For example, *hierarchically well separated trees* (see [4, 8]) satisfy Properties 1–3 and Property 4 on expectation.

We use tree T' to define a *Shapley cost share* for each terminal pair in $S(T)$. Let $T'[S_i(T)]$ be the induced subtree of T' on terminals pair set $S_i(T)$. For a terminal pair $(s_i, t_i) \in S(T)$, we define $\xi'_i(S_i(T))$ to be the sum of the respective Shapley cost shares of terminals s_i and t_i in $T'[S_i(T)]$. The following lemma follows immediately from the definition of Shapley cost shares.

Lemma 6.2. *Let ξ' be the Shapley cost shares of terminal pairs in $S(T)$. Then*

$$\sum_{(s_i,t_i)\in S(T)} \xi'_i(S_i(T)) \leq H_k \cdot \ell(T').$$

We next show that the cost share $\xi_i(S_i(T))$ of terminal pair (s_i, t_i) is upper bounded by its corresponding Shapley cost share $\xi'_i(S_i(T))$ in $T'[S_i(T)]$. This together with the above lemma and Property 4 shows $O(\log^2 k)$-summability of ξ for identical death times.

Lemma 6.3. *The cost share $\xi_i(S_i(T))$ of terminal pair $(s_i, t_i) \in S(T)$ is at most its Shapley cost share $\xi'_i(S_i(T))$.*

Proof. All terminals in $S(T)$ are active until time D. The cost share $\xi_u(S_i(T))$ of a terminal $u \in \{s_i, t_i\}$ in KLS is then defined as

$$\xi_u(S_i(T)) = \int_{\tau=0}^{D} \frac{d\tau}{a_i^\tau(u)}$$

where $a_i^\tau(u)$ is the number of active terminals in u's moat at time τ in the run of KLS$(S_i(T))$. We bound the cost share that $u = s_i$ receives in KLS$(S_i(T))$ by its Shapley cost share. An analogous argument holds for $u = t_i$.

Consider the induced subtree $T'_i = T'[S_i(T)]$ on $S_i(T)$. Let $P_{ur} = (e_1, \ldots, e_m)$ be the unique u, r-path in T'_i. Consider an edge e_j, $1 < j \leq m$ and let $T'_i(e_j)$ be the subtree of T'_i below edge e_j. We use $m_i(e_j)$ to refer to the number of terminals in $T'_i(e_j)$; define $m_i(e_1) = 1$. The Shapley cost share that u received for edge e_j is $\ell(e_j)/m_i(e_j)$. Thus,

$$\xi'_u(S_i(T)) = \sum_{j=1}^{m} \frac{\ell(e_j)}{m_i(e_j)}.$$

Let x be any terminal in $T'_i(e_j)$. By Property 2, we have $d_G(u, x) \leq \ell(e_j)$. Since both x and u are active until time D, their respective moats in KLS$(S_i(T))$ must have met by time at most $d_G(u, x)/2 \leq \ell(e_j)/2 = \ell(e_{j-1})$. Thus, $a_i^\tau(u) \geq m_i(e_j)$ for all $\tau \geq \ell(e_{j-1})$ for all $1 < j \leq m$.

Note that the cost share that u receives up to time 1 is at most 1. As $\ell(e_1) = 1$ and $\ell(e_m) \geq D$, we can write

$$\begin{aligned}
\xi_u(S_i(T)) = \int_{\tau=0}^{D} \frac{d\tau}{a_i^\tau(u)} &\leq 1 + \sum_{j=2}^{m} \int_{\tau=\ell(e_{j-1})}^{\ell(e_j)} \frac{d\tau}{a_i^\tau(u)} \\
&\leq 1 + \sum_{j=2}^{m} \int_{\tau=\ell(e_{j-1})}^{\ell(e_j)} \frac{d\tau}{m_i(e_j)} \\
&= 1 + \sum_{j=2}^{m} \frac{\ell(e_{j-1})}{m_i(e_j)} \leq \xi'_u(S_i(T)).
\end{aligned}$$

References

[1] Ajit Agrawal, Philip Klein, and R. Ravi. When trees collide: an approximation algorithm for the generalized Steiner problem on networks. *SIAM J. Comput.*, 24(3):440–456, 1995. (Preliminary version in *23rd STOC*, 1991).

[2] Aaron Archer, Joan Feigenbaum, Arvind Krishnamurthy, Rahul Sami, and Scott Shenker. Approximation and collusion in multicast cost sharing. *Games and Economic Behavior*, 47(1):36–71, 2004.

[3] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, 1998.

[4] Yair Bartal. Probabilistic approximations of metric spaces and its algorithmic applications. In *Proceedings of the 37th Symposium on the Foundations of Computer Science (FOCS)*, pages 184–193, 1996.

[5] M. Bern and P. Plassman. The Steiner problem with edge lengths 1 and 2. *Information Processing Letters*, 32:171–176, 1989.

[6] Daniel Bienstock, Michel X. Goemans, David Simchi-Levi, and David P. Williamson. A note on the prize collecting traveling salesman problem. *Math. Program.*, 59:413–420, 1993.

[7] S. Chawla, T. Roughgarden, and M. Sundararajan. Optimal cost-sharing mechanisms for steiner forest problems. submitted to WINE.

[8] Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. *J. Comput. System Sci.*, 69(3):485–497, 2004.

[9] Joan Feigenbaum, Arvind Krishnamurthy, Rahul Sami, and Scott Shenker. Hardness results for multicast cost-sharing. *Theoretical Computer Science*, 304:215–236, 2003.

[10] Joan Feigenbaum, Christos H. Papadimitriou, and Scott Shenker. Sharing the cost of multicast transmissions. *J. Comput. System Sci.*, 63(1):21–41, 2001. Special issue on internet algorithms.

[11] Michel X. Goemans and David P. Williamson. A general approximation technique for constrained forest problems. *SIAM J. Comput.*, 24(2):296–317, 1995. (Preliminary version in *5th SODA*, 1994).

[12] J. Green, E. Kohlberg, and J. J. Laffont. Partial equilibrium approach to the free rider problem. *Journal of Public Economics*, 6:375–394, 1976.

[13] A. Gupta, A. Srinivasan, and É. Tardos. Cost-sharing mechanisms for network design. In *Proceedings of the Seventh International Workshop on Approximation Algorithms for Combinatorial Optimization Problems*, 2004.

[14] Mohammad Taghi Hajiaghayi and Kamal Jain. The prize-collecting generalized steiner tree problem via a new approach of primal-dual schema. In *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2006)*, pages 631–640, 2006.

[15] N. Immorlica, M. Mahdian, and V. S. Mirrokni. Limitations of cross-monotonic cost sharing schemes. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 602–611. ACM Press, 2005.

[16] Kamal Jain and Vijay Vazirani. Applications of approximation algorithms to cooperative games. In *Proceedings of the 33rd Annual ACM Symposium on the Theory of Computing (STOC)*, pages 364–372, 2001.

[17] Kathryn J. Kent and Darko Skorin-Kapov. Population monotonic cost allocations on MSTs. In *Proceedings of the 6th International Conference on Operational Research (Rovinj, 1996)*, pages 43–48. Croatian Oper. Res. Soc., Zagreb, 1996.

[18] Jochen Könemann, Stefano Leonardi, and Guido Schäfer. A group-strategyproof mechanism for Steiner forests. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 612–619. ACM Press, 2005.

[19] Jochen Könemann, Stefano Leonardi, Guido Schäfer, and Stefan van Zwam. From primal-dual to cost shares and back: a stronger LP relaxation for the Steiner forest problem. In *Automata, languages and programming*, volume 3580 of *Lecture Notes in Comput. Sci.*, pages 930–942. Springer, Berlin, 2005.

[20] Stefano Leonardi and Guido Schäfer. Cross-monotonic cost sharing methods for connected facility location games. *Theor. Comput. Sci.*, 326(1-3):431–442, 2004.

[21] Hervé Moulin. Incremental cost sharing: Characterization by coalition strategy-proofness. *Social Choice and Welfare*, 16:279–320, 1999.

[22] Hervé Moulin and Scott Shenker. Strategyproof sharing of submodular costs: budget balance versus efficiency. *Econom. Theory*, 18(3):511–533, 2001.

[23] Martin Pál and Éva Tardos. Group strategyproof mechanisms via primal-dual algorithms. In *Proceedings of the 44th Symposium on the Foundations of Computer Science (FOCS)*, pages 584–593, 2003.

[24] K. Roberts. The characterization of implementable choice rules. In J. J. Laffont, editor, *Aggregation and Revelation of Preferences*. North-Holland, 1979.

[25] T. Roughgarden and M. Sundararajan. New trade-offs in cost-sharing mechanisms. In *STOC*, 2006.