Bounding the payment of approximate truthful mechanisms

Gruia Calinescu[1]

Department of Computer Science, Illinois Institute of Technology, Chicago, IL 60616.
E-mail: calinesc@iit.edu.

Abstract. In a STACS 2003 paper, Talwar analyses the overpayment the VCG mechanism incurs for ensuring truthfulness in auction. Among other results, he studies k-Set Cover (given a universe U and a collection of sets $S_1, S_2, \ldots, S_q$, each having a cost $c(S_i)$ and at most k elements of U, find a minimum cost subcollection, called cover, whose union equals U) and shows that the payment of the optimum cover OPT is at most $kc(OPT')$, where OPT' is the best cover disjoint from the optimum cover. For $k \geq 3$, k-Set Cover is known to be NP-Hard, and thus truthful mechanisms based on approximation algorithms are desirable. We show that the payment incurred by two approximation algorithms (including the Greedy algorithm) is bounded by $(k-1)c(OPT) + kc(OPT')$. The same approximation algorithms have payment bounded by $k(c(OPT) + c(OPT'))$ when applied to more general set systems, which include k-Polymatroid Cover, a problem with applications in Steiner Tree computations. If q is such that an element in a k-Set-Cover instance appears in at most q sets, we show that the payment of our algorithms is bounded by qk^2 times the payment of the optimum algorithm.

1 Introduction

There has been a surge of recent interest in the intersection area of economic sciences and computer science (see [21] for a survey). Inside this area is the field of *mechanism design*, which we describe below in the particular setting we are going to analyse in this paper. The seminal paper of Nisan and Ronen [20] introduced the computational issues of mechanism design and gives a general overview in a setting more general than ours. A reader with economics background will recognize the scenario in our simplified case. A reader with computer science background will find below all the definitions used in this paper, and is warned that more definitions and notations are required in the general setting.

We are given a set E of elements and a family $\mathcal{F}$ of *feasible* subsets of E. We assume the set system $(E, \mathcal{F})$ is closed upwards, which means that for any $T \in \mathcal{F}$ and any superset $T' : T \subseteq T' \subseteq E$, we have $T' \in \mathcal{F}$. Each element $e \in E$ is controlled by a different economic agent, which we called the agent of e. The agent of e has a private (unknown to anyone else) cost $c(e) \geq 0$ for providing e to a feasible subset. Finding feasible subsets is sometimes called *team selection* in the literature [3, 11] dealing with mechanisms.

A *direct revelation mechanism* is a protocol which asks each agent to provide a bid $b(e)$, and then computes a feasible subset $M = M(b)$ and a payment function $p(e) \geq 0$. If $e \in M$, the agent of e must allow e to be used by M in exchange for the payment $p(e)$.

The agents are selfish and could misrepresent their cost in order to increase their payments. The field of *mechanism design* deals with the design of protocols which ensure the designer's goal are achieved by giving incentives to agents. The revelation principle [24, 18] states that in order to prove or disprove the existence of good mechanism for a given problem, one can consider only truthful revelation mechanism. Thus we only use direct revelation mechanisms, which are characterized by the computation of the feasible subset M and the payment function p.

In this context, the economics notion of maximizing social welfare is minimizing the cost of M. Minimizing the cost of M could be the goal of the designer of the mechanism. Another reasonable goal is minimizing the sum of payments. A mechanism is called *truthful* (also called strategy-proof) if every agent has interest to bid her cost (by setting $b(e) = c(e)$) regardless of the other bidders' bids.

Consider the simple case in which each element is a feasible subset, a situation which occurs frequently in real life. Say the government needs a task to be performed and invites sealed bids from agents (contractors). If the government chooses the lowest bidder and pays to this agent her bid, then the agents might bid higher then their cost with the goal of making a profit. The government can however use the *VCG mechanism* [26, 9, 12]: select the lowest bidder and pay her an amount equal to the second lowest bid. Assuming agent don't collude, it can be shown that in this case each agent has the interest to bid her cost. The VCG mechanism is truthful, and minimizes cost (maximizes social welfare). However the payment may be much larger than the cost.

The VCG mechanism can be applied to the set system setting as well [25, 3, 6], and selfish agent would bid their cost: for all e, we have $b(e) = c(e)$. The mechanism selects OPT, a feasible subset of minimum cost, and computes payments with a formula we present later.

The set systems we consider are instances of problems. For example, the elements can be the edges of a graph and feasible subsets consists of connected spanning subgraphs: the Minimum Spanning Tree problem. Of particular interest in routing protocols is the Shortest s-t Path problem [3, 11].

Let $p(I, \mathcal{A})$ be the payment of mechanism $\mathcal{A}$ on instance I, that is the sum of the payments given by the mechanism to the agents. Talwar [25] analyses the *frugality ratio* of problems, which is defined as the maximum over instances I of $p(I, VCG)/c(OPT')$, where OPT' is the best feasible subset disjoint from OPT. [25] characterizes the problems with frugality ratio 1. For example, Minimum Spanning Tree has frugality ratio 1 [6, 25], and Shortest s-t Path has frugality ratio $\Theta(n)$ [3, 25], where n is the number of vertices in the graph.

One interesting problem analysed by Talwar [25] is the k-Set Cover problem: given a universe U and a collection of sets $S_1, S_2, \ldots, S_q$, each having a cost $c(S_i)$

and at most k elements of U, find a minimum cost subcollection (called *cover*) whose union equals U. k-Set Cover fits the general framework as follows: each set S_i is an element of E, and covers are the feasible subsets. [25] has proved that the frugality ratio of k-Set Cover is exactly k. Thus for k-Set Cover, VCG, besides optimizing the cost (social welfare), guarantees a bound on the payment.

For $k \geq 3$, k-Set Cover is known to be NP-Hard, and thus truthful mechanisms based on approximation algorithms are desirable. It is mentioned in [3, 1] that it is known that a mechanisms is truthful if it is *monotone* (defined later). For k-Set Cover, we analyze the monotone mechanism based on the Greedy (G) approximation algorithms. We show that its payment is bounded by $(k-1)c(OPT)+kc(OPT')$. This bound is at most twice worse than the bound of the VCG mechanism, which must find the optimum solution. We present an instance of k-Set Cover where the payment of Greedy is lower than the payment of OPT.

If q is such that an element in a k-Set-Cover instance appears in at most q sets, then we can prove the bound on the payments: $p(I,G) \leq k^2 q \cdot p(I, VCG)$. We call (k,q)-Set Cover the Set Cover problem where each set has size at most k and every element appears in at most q sets. (k,q)-Set Cover is the only problem we are aware of where there is a nontrivial bound on the ratio of the payment of an approximation algorithm to the payment of the exact algorithm in our particular setting. Previous approximate truthful mechanisms [1, 2] either are for maximization problems and look at revenue (instead of payment) or look at a different setting in which costs are public and the agent have some other private data.

In the k-Polymatroid Cover problem, the set system is defined by a a submodular rank function $r : \mathcal{P}(E) \to \mathbb{N}$ with $r(e) \leq k$ for all $e \in E$: a feasible set T is a set satisfying $r(T) = r(E)$. The k-Polymatroid Cover problem has application for Steiner Tree computation [22, 5, 23, 28, 15, 14], and is a generalization of k-Set Cover. The k-Polymatroid problem is NP-Hard for $k \geq 3$, polynomial for $k = 1$ (when the problem becomes finding a minimum-cost basis in a matroid), while the complexity for $k = 2$ is unknown, but pseudopolynomial algorithms are known for linear polymatroids [7, 22]. Wolsey [27] has shown that Greedy has approximation ratio H_k for k-Polymatroid Cover, where $H_k = 1 + \frac{1}{2} + \cdots + \frac{1}{k}$ is the kth harmonic number. We show that $p(I,G) \leq k\,(c(OPT) + c(OPT'))$ and use techniques from [25] to show that $p(I, VCG) \leq kc(OPT')$.

A second algorithm applies any set system with an oracle deciding whether a set is feasible or not. Space limitations prevent us from discussing this algorithm, and we just mention that we can prove exactly the same bounds on its payment as we do for Greedy when the set system is given by k-Set Cover, (k,q)-Set Cover, or k-Polymatroid Cover.

The paper is organized as follows. Section 2 gives the definition of monotone mechanisms with some known associated properties, definitions related to k-Polymatroids, and the definitions of the Greedy algorithm, with some associated properties. Section 3 compares the payment of Greedy to $c(OPT) +$

$c(OPT')$. Section 4 compares the payment of the approximation algorithm directly to the payment of the optimum algorithm.

2 Preliminaries

We start by formally defining truthfulness. Let $M(b)$ be the feasible subset computed by the mechanism given bid vector $b = b(e) \mid e \in E$, where E is the set of elements. For agent e and bid vector b, we define $profit_e(b) = p(e) - c(e)$, if $e \in M(b)$ and $profit_e(b) = 0$ otherwise. We assume that each agent's goal is to maximize her profit. The algorithm which computes $M(b)$ is known to all before the bidding, and the payments $p(e)$ are known to all after the bidding. We require that our mechanisms satisfy the *voluntary participation* condition, which states that the profit of an agent cannot be negative (or the agent might decide not to participate).

Given a mechanisms as described above, let b_{-e} denote the vector of all bids except the one made by the agent of e, so we can write $b = (b_{-e}, b(e))$. We say that *truthtelling* is a *(weakly) dominant strategy* for the agent of e if, regardless of b_{-e}, bidding $b(e) = c(e)$ maximizes $profit_e(b)$. So, even if the agent of e knew the bids of the other agents ahead of time, the best she could do is bid the truth cost. If truthtelling is a dominant strategy for every agent we call the mechanism *truthful*.

2.1 Monotone mechanisms

Let $b(e)$ be the bid values of the agents, and let $M(b)$ be the subset picked by the mechanism, which depends on the bids b. The mechanism is *monotone* if for every element $e \in M(b)$ a lower bid $\overline{b}(e) < b(e)$ only by the agent of e (that is, $\overline{b}(f) = b(f)$ for $f \in E \setminus \{e\}$), also result in $e \in M(\overline{b})$. In this case there is a threshold $t(e) = t(e, b_{-e})$ for every $e \in E$ which depends on the algorithms which picks M and b_{-e}, such that:

- If $b(e) < t(e, b_{-e})$, then $e \in M(b)$
- If $b(e) > t(e, b_{-e})$, then $e \notin M(b)$

One defines

$$p(e,b) = \begin{cases} 0 & \text{if } e \notin M(b) \\ t(e, b_{-e}) & \text{if } e \in M(b) \end{cases}$$

If $M(b)$ is an optimum solution, then this mechanism is monotone, and in fact coincides with the VCG mechanism, which is defined as follows. Let $b[e \to x]$ denote the cost function $\overline{b}$ which is equal to b everywhere except that $\overline{b}(e) = x$. Let $opt(b) = b(OPT)$, the total bid-value of the optimum solution for bids b. Then:

$$p(e,b) = \begin{cases} 0 & \text{if } e \notin OPT(b) \\ opt([b \to \infty]) - opt([b \to 0]) & \text{if } e \in OPT(b) \end{cases}$$

For the sake of completeness we include a proof of the folklore [3] result:

Theorem 1. *A monotone mechanism is truthful.*

Proof. First we note that since $p(e, b) \geq b(e)$ whenever $e \in M$, and $p(e, b) = 0$ whenever $e \notin M(b)$ (this is the case when no services or money is traded), the agent has an interest to participate in the mechanism. Why would an agent bid other than her cost? If the agent of e bids $b(e) < c(e)$, then three outcomes are possible:

- $e \notin M(b)$. The agent of e does not trade services or money, the same outcome as in the case when $b(e) = c(e)$ (the monotonicity of the mechanism implies that $e \notin M$ in when $b(e) = c(e)$ as well)
- $e \in M(b)$ and $c(e) > t(e, b_{-e})$. In this case the agent of e is losing money (as $p(e, b) = t(e, b_{-e})$), while bidding $b(e) = c(e)$ would result in no trade or a trade with payment being equal to the cost.
- $e \in M(b)$ and $c(e) \leq t(e, b_{-e})$. In this case the agent receives the same payment as in the case when she bids $b(e) = c(e)$

On the other side, if the agent of e bids $b(e) > c(e)$, then again three outcomes are possible:

- $e \in M(b)$. As $t(e, b_{-e}) \geq b(e)$, the payment received by the agent of e is the same as in the case when it bids $b(e) = c(e)$ (the monotonicity of the mechanism implies that $e \in M$ in this case as well)
- $e \notin M(b)$ and $c(e) < t(e, b_{-e})$. In this case the agent does not trade, while bidding $b(e) = c(e)$ would have resulted in a payment exceeding the cost
- $e \notin M(b)$ and $c(e) \geq t(e, b_{-e})$. In this case the agent does not trade. Bidding $b(e) = c(e)$ would have resulted also in a no-trade, or in a trade with payment being equal to the cost.

In conclusion, in no case can the agent benefit from not bidding her cost. ■

Xiang-Yang Li and Weizhao Wang [16] have implicitly proved that the Greedy algorithm when applied to Set Cover is monotone.

2.2 Polymatroids and approximation algorithms

A rank function $r : \mathcal{P}(E) \to \mathbb{N}$ is called submodular if for any two sets $A, B \subseteq E$

$$r(A \cap B) + r(A \cup B) \leq r(A) + r(B)$$

A k-polymatroid is given by a submodular rank function satisfying $r(\{e\}) \leq k$ for all $e \in E$. The k-Polymatroid Cover problems asks for a minimum cost subset of E with rank equal to $r(E)$. k-Polymatroid Cover is a generalization of k-Set Cover by setting for a collection of sets A the rank $r(A) = | \cup_{S \in A} S|$. The Greedy algorithm for Set Cover analysed by Chvatal [8] can be used for k-Polymatroid Cover [27] as follows: let A be the set of elements picked so far; initially $A = \emptyset$. If $r(A) < r(E)$, find $g \in E$ such that $\frac{r(A \cup \{g\}) - r(A)}{c(g)}$ is maximized, and set $A \leftarrow A \cup \{g\}$. We generalize and simplify [16] to obtain:

Theorem 2. GREEDY *for k-Polymatroid Cover is monotone.*

Proof. Assume we decrease the bid of e from $b(e)$ to $\overline{b}(e)$, while for the other elements $f \in E$ we keep $\overline{b}(f) = b(f)$. The GREEDY algorithm proceeds as before the decrease and picks e, unless e is picked sooner. ■

Since we only use monotone algorithms in this paper, and truthful mechanisms based on them, from now on we do not mention the bids, as we assume the bids equal the costs.

3 The GREEDY Algorithm

In this section we compare the payment of the GREEDY algorithm to $c(OPT')$. We start with k-Set Cover, where some arguments of Talwar [25] help.

Theorem 3. *Assume that for a k-Set-Cover instance with collection of sets T, there is a cover in $T \setminus OPT$. Let OPT' be the minimum cost cover in $T \setminus OPT$. Then $p(G) \leq (k-1)c(OPT) + kc(OPT')$, where $p(G)$ is the payment of the* GREEDY *algorithm.*

Proof. Let $T = S_1, S_2, \ldots, S_n$ be the collection of sets and let $B_1, B_2, \ldots, B_m$ be the sets picked by GREEDY in this order. Let Q_i be the subset of B_i with the elements first covered by B_i (that is, not covered by $B_1, B_2, \ldots, B_{j-1}$), and note that for $1 \leq i < j \leq m$, we have $Q_i \cap Q_j = \emptyset$.

Consider a set B_i and we analyze $p(B_i)$. We increase $c(B_i)$ until GREEDY does not pick B_i. We reach the following situation: GREEDY picks $B_1, B_2, \ldots, B_{i-1}$, then it might pick a several sets H_j, and then it picks B_i. However, a further increase in $c(B_i)$ results in B_i not being picked. Let $\overline{c}$ be this new cost function which differs from c only for B_i, and note that $p(B_i) = \overline{c}(B_i)$. Define $\overline{Q}_i$ to be the elements first covered by B_i in the cover given by GREEDY when run with $\overline{c}$, and note that $\overline{Q}_i \subseteq Q_i$.

Assume first that $B_i \notin OPT'$. We know $\overline{Q}_i$ is covered by sets $R_1, R_2, \ldots, R_a$ of OPT', and note that GREEDY did not pick any of R_j before B_i when run on cost $\overline{c}$. By the way GREEDY operates, for $j = 1, 2, \ldots, a$:

$$c(R_j) \geq \frac{|R_j \cap \overline{Q}_i|}{|\overline{Q}_i|} \overline{c}(B_i). \tag{1}$$

Summing up over j, we obtain that $\overline{c}(B_i) \leq \sum_{j=1}^{a} c(R_j)$, and we say that B_i charges each R_j. Since $|R_j| \leq k$ and the sets $\overline{Q}_i$ are disjoint, it follows that a set R in OPT' can be charged at most k times. We have to do more work since when $B_i \in OPT'$, we cannot always find sets in OPT' to be charged as above by B_i.

Now assume that $B_i \in OPT'$. Using the argument as above, we find several $O_j \in OPT$ such that B_i charges O_j, and O_j is charged in total at most k times. At this moment we have proven $p(G) \leq k(c(OPT') + c(OPT))$.

If O_j from OPT is charged exactly k times, then it is charged by k sets from OPT'. Each of these k sets of OPT' is using a different element to charge O_j and therefore they together cover O_j. Since O_j can be replaced by these k sets of OPT', their total cost is at least $c(O_j)$. We move the charges as follows: each of the k sets B_i of OPT' which charges O_j is asked to charge only $c(O_j) - c(B_i)$; by the argument above this reduces the charge of O_j to at most $(k-1)c(O_j)$. B_i is charging $c(B_i)$ to itself. Now B_i can be charged by other sets of G, but it can be charged at most $k-1$ times since it cannot be charged through the element it uses to charge O_j, as that element is in $\overline{Q}_i$ and not in any other $\overline{Q}_r$, with B_r in G. Thus B_i is charged in total at most k times.

With the modified charges, each set in OPT is charged at most $k-1$ times and each set of OPT' is charged at most k times. Thus we have $p(G) \leq (k-1)c(OPT) + kc(OPT')$. ■

We do not have a tight example for the bound in Theorem 3 and suspect the bound is not tight. We include an example (see also Figure 1) taken from [25] which shows $p(OPT) = k \cdot c(OPT')$ and for which $p(G) = k \cdot c(OPT')$ too. There are $k+1$ sets $S_0, S_1, \ldots, S_k$ of costs $c(S_0) = 1$ and $c(S_i) = 0$ for $i = 1, 2, \ldots, k$. All the k elements $e_1, e_2, \ldots, e_k$ are included in S_0, while $S_i = \{e_i\}$. GREEDY (and also OPT) picks the 0-cost sets and one can check it pays 1 for each of them.

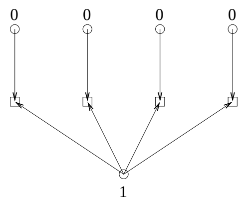

Fig. 1. An example where the payment $p(OPT) = p(G) = k \cdot c(OPT')$ for $k = 4$. The sets are represented by circles and the elements by squares. GREEDY (and also OPT) picks the sets of cost 0 and pays 1 for each of them, while $c(OPT') = 1$.

For the more general k-Polymatroid, we obtain a slightly weaker bound.

Theorem 4. *Assume the set system $(E, \mathcal{F})$ is given by a k-Polymatroid, and that for given cost function c, there is a feasible subset in $E \setminus OPT$. Let OPT' be the minimum cost feasible subset in $E \setminus OPT$. Then $p(G) \leq k\,(c(OPT) + c(OPT'))$, where $p(G)$ is the payment of the* GREEDY *algorithm.*

Proof. Some of the arguments below are taken from the analysis of the GREEDY algorithm given in [4]. Helgason [13], McDiamid [19], and Lovasz [17] have shown that each polymatroid $(E, \mathcal{F})$ can be obtained from a matroid M with rank function $\overline{r}$ by the following construction: An element $h \in E$ corresponds to a set $M_h \in M$ such that for all $X \subseteq E$ we have $r(X) = \overline{r}(\cup_{h \in X} M_h)$.

Let $n := r(E) = \overline{r}(M)$ and let $g_1, g_2, \ldots, g_m$ be, in this order, the elements picked by GREEDY. Let $B_i = M_{g_i}$, as defined above. Let $S_0 := A_0 := \emptyset$, and for $i = 1, 2, \ldots, m$, let $A_i \subseteq M$ be defined by $A_i = \cup_{j=1}^{i} B_i$. For $i = 1, 2, \ldots, m$, let $Q_i \subseteq B_i$ be such that $S_i := \cup_{j=1}^{i} Q_j$ is a basis for A_i.

Consider an element g_i picked by GREEDY and we analyze $p(g_i)$. We increase $c(g_i)$ until GREEDY does not pick g_i. We reach the following situation: GREEDY picks $g_1, g_2, \ldots, g_{i-1}$, then it might pick a set of elements H, and then it picks g_i. However, a further increase in $c(g_i)$ results in g_i not being picked. Let $\overline{c}$ be this new cost function which differs from c only for g_i, Define $Z_i \subseteq \cup_{h \in H} M_h$ such that $S_{i-1} \cup Z_i$ is a basis for $A_{i-1} \cup (\cup_{h \in H} M_h)$. Now define $\overline{Q}_i$ such that $\overline{Q}_i \subseteq Q_i$ and $S_{i-1} \cup Z_i \cup \overline{Q}_i$ is a basis in $A_{i-1} \cup (\cup_{h \in H} M_h) \cup Q_i$, which means $S_{i-1} \cup Z_i \cup \overline{Q}_i$ is also a basis in $A_{i-1} \cup (\cup_{h \in H} M_h) \cup B_i$ since $S_{i-1} \cup Q_i$ spans $A_{i-1} \cup B_i$.

Let $T \subseteq \cup_{h \in OPT'} M_h$ be another basis of M. For every $f \in T$ choose a $\psi(f) \in OPT'$ such that $f \in M_{\psi(f)}$. For all $h \in OPT'$, the set $\psi^{-1}(h)$ is independent and therefore:

$$|\psi^{-1}(h)| \leq r(h) \leq k \tag{2}$$

In order to analyse the payment of GREEDY, we need to select from T disjoint sets $\overline{X}_i$, with $i = 1, 2, \ldots, m$, such that $|\overline{X}_i| = |\overline{Q}_i|$ and each $A_{i-1} \cup Z_i \cup \overline{X}_i$ is independent. We ensure disjointness by constructing a partition of T into sets X_i, with $i = 1, 2, \ldots, m$ such that $\overline{X}_i \subseteq X_i$. Borrowing again from [4], we construct $\overline{X}_i$ and X_i one by one, but in a more involved procedure described below. Let $T_0 := T$ and $T_i := T \setminus \left(\cup_{j=1}^{i} X_i\right)$. We impose the invariant that for all $i = 0, 1, \ldots, m$, $S_i \cup T_i$ forms a base in M.

We process $i = 1, 2, \ldots, m$ as follows. We complete $S_{i-1} \cup Z_i \cup \overline{Q}_i$ to a base using a set $W_i \subseteq T_{i-1}$; such a completion exists since $S_{i-1} \cup T_{i-1}$ forms a base. We contract $S_{i-1} \cup W_i$, obtaining a new matroid M_i, in which we have two bases: $Z_i \cup \overline{Q}_i$ and $T_{i-1} \setminus W_i$. Using Exercise 8.48 from [10] or its generalization in Lemma 1 below, $T_{i-1} \setminus W_i$ can be partition into two sets, one of them we call $\overline{X}_i$, such that $(T_{i-1} \setminus W_i \setminus \overline{X}_i) \cup \overline{Q}_i$ and $Z_i \cup \overline{X}_i$ are both bases in the matroid M_i. Thus $S_{i-1} \cup Z_i \cup \overline{X}_i$ is independent and $S_{i-1} \cup \overline{Q}_i \cup \left(T_{i-1} \setminus \overline{X}_i\right)$ is a base in M.

Now we can use the procedure described in [4] to get X_i such that $S_i \cup (T_{i-1} \setminus X_i)$ is a base. If we denote $F_i := Q_i \setminus \overline{Q}_i$, we have $S_i = S_{i-1} \cup Q_i = S_{i-1} \cup \overline{Q}_i \cup F_i$. If $F_i = \emptyset$, we set $X_i = \overline{X}_i$ and we can proceed to $i+1$. Else, let $f_1, \ldots, f_r$ be the elements of F_i. Starting with $j = 1$, we find elements $t_j \in T_{i-1} \setminus \overline{X}_i \setminus \{t_1, t_2, \ldots, t_{j-1}\}$ such that $S_{i-1} \cup \overline{Q}_i \cup \{f_1, f_2, \ldots, f_j\} \cup \left(T_{i-1} \setminus \overline{X}_i \setminus \{t_1, t_2, \ldots, t_j\}\right)$ is a base. The element t_j is found as follows: If $f_j \in \left(T_{i-1} \setminus \overline{X}_i \setminus \{t_1, t_2, \ldots, t_{j-1}\}\right)$, then $t_j = f_j$. Else, adding f_j to the base $S_{i-1} \cup \overline{Q}_i \cup \{f_1, f_2, \ldots, f_{j-1}\} \cup \left(T_{i-1} \setminus \overline{X}_i \setminus \{t_1, t_2, \ldots, t_{j-1}\}\right)$ creates a circuit C. As f_j is not spanned by $S_{i-1} \cup \overline{Q}_i \cup \{f_1, f_2, \ldots, f_{j-1}\}$, we can find $t_j \in C \setminus \left(S_{i-1} \cup \overline{Q}_i \cup \{f_1, f_2, \ldots, f_{j-1}\}\right)$. It follows that $S_{i-1} \cup \overline{Q}_i \cup \{f_1, f_2, \ldots, f_{j-1}\} \cup \left(T_{i-1} \setminus \overline{X}_i \setminus \{t_1, t_2, \ldots, t_{j-1}\}\right) \setminus \{t_j\} \cup \{f_j\}$ is also a base. Then we set $X_i = \overline{X}_i \cup \{t_1, t_2, \ldots, t_j\}$ and we note that indeed $S_i \cup T_i$ forms a base thus maintaining the invariant.

Thus we selected from T disjoint sets $\overline{X}_i$, with $i = 1, 2, \ldots, m$, such that $|\overline{X}_i| = |\overline{Q}_i|$ and each $A_{i-1} \cup Z_i \cup \overline{X}_i$ is independent.

Assume first that $g_i \notin OPT'$. We know B_i increases the rank of $S_{i-1} \cup Z_i$ by $|\overline{Q}_i|$. $\overline{X}_i$ also increases the rank of $S_{i-1} \cup Z_i$ by $|\overline{X}_i| = |\overline{Q}_i|$. Let $h_1, h_2, \ldots, h_a$ be elements of OPT' such that $\psi(h_i) \cap \overline{X}_i \neq \emptyset$, and note that GREEDY did not pick any of h_j before g_i when run on cost $\overline{c}$. Define $r_j = r\left(S_{i-1} \cup Z_i \cup M_{h_j}\right) - r\left(S_{i-1} \cup Z_i\right)$. By the way GREEDY operates, for $j = 1, 2, \ldots, a$:

$$c(h_j) \geq \frac{r_j}{|\overline{Q}_i|} \overline{c}(g_i). \tag{3}$$

Summing up over j and using the fact (obtained from submodularity) that $\sum_{j=1}^{a} r_j \geq r\left(S_{i-1} \cup Z_i \cup \overline{X}_i\right) - r\left(S_{i-1} \cup Z_i\right) = |\overline{X}_i| = |\overline{Q}_i|$, we obtain that $\overline{c}(g_i) \leq \sum_{j=1}^{a} c(h_j)$, and we say that g_i charges each h_j. From Equation 2 and the fact that the sets $\overline{X}_i$ are disjoint we obtain that each element of OPT' can be charged at most k times.

Now assume that $g_i \in OPT'$. Using the same argument, but with the basis T derived from OPT rather than OPT', we charge the payment of g_i to the cost of one or more elements of OPT, and such that each element of OPT can be charged at most k times. At this moment we have proven $p(G) \leq k(c(OPT') + c(OPT))$. ■

The following property of matroids generalizes Exercise 8.48 of [10]. Due to lack of space we omit the proof of the lemma and the following theorem, which uses Talwar's technique and the lemma.

Lemma 1. *Let M be a matroid on set N and A, B bases of M. Let $X_1, X_2, \ldots, X_q$ be a partition of A. Then there is a partition $Y_1, Y_2, \ldots, Y_q$ of B such that for all $i = 1, 2, \ldots, q$, $A \setminus X_i \cup Y_i$ is a base.*

Theorem 5. *Assume the set system $(E, \mathcal{F})$ is given by a k-Polymatroid, and that for cost function c, there is a feasible subset in $E \setminus OPT$. Let OPT' be the minimum cost feasible subset in $E \setminus OPT$. Then $p(OPT) \leq kc(OPT')$.*

We include a k-Set-Cover example that show that the payment of GREEDY can be lower than the payment of OPT. We use $k = 2$; note that this problem, also known as Edge Cover, has polynomial-time algorithms. An illustration of the example is given in Figure 2. OPT consists of three sets: $\{1, 2\}$, $\{3, 4\}$, and $\{5, 6\}$ each of cost $1 + \epsilon$. OPT', which is also G, consist of four sets each of cost 1: $\{1, 3\}$, $\{2, 3\}$, $\{4, 5\}$, and $\{4.6\}$. One can check that $p(OPT) = 3(2 - 2\epsilon)$ while $p(G) = 4(1 + \epsilon)$.

4 Bounds in terms of $p(OPT)$

In this section we analyse the payment of the approximation algorithms versus the payment of the optimum algorithm. For k-Set-Cover, the payment of GREEDY can be much higher than the payment of OPT, as shown by the following example, also illustrated in Figure 3. We have $1 + q(k - 1)$ elements

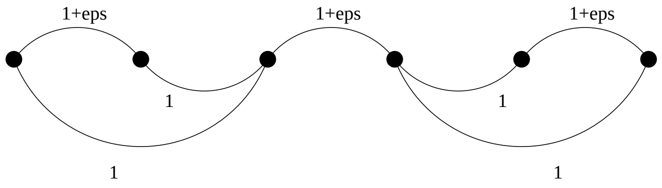

Fig. 2. An example where the payment of OPT is larger than the payment of the approximation algorithm GREEDY. The sets are the edges of this graph.

$d_0, d_1, \ldots, d_{q(k-1)}$. Set $S_0 = \{d_0\}$ and $c(S_0) = 1$. For $i = 1, 2, \ldots, q(k-1)$, set $S_i = \{d_i\}$ and $c(S_i) = 0$. For $j = 1, 2, \ldots, q$, the set $B_j = \{d_0\} \cup \{d_{(j-1)(k-1)+1}, d_{(j-1)(k-1)+2}, \ldots, d_{j(k-1)}\}$, each of cost $1 + \epsilon$. One can check $p(OPT) = 1 + \epsilon(1 + q(k-1))$. GREEDY also finds the optimum solution, and one can check that for each $i = 1, 2, \ldots, q(k-1)$, GREEDY pays $\frac{1+\epsilon}{2}$. Thus $p(G) \geq (\frac{1}{2} - \epsilon)(k-1)qp(OPT)$ is possible.

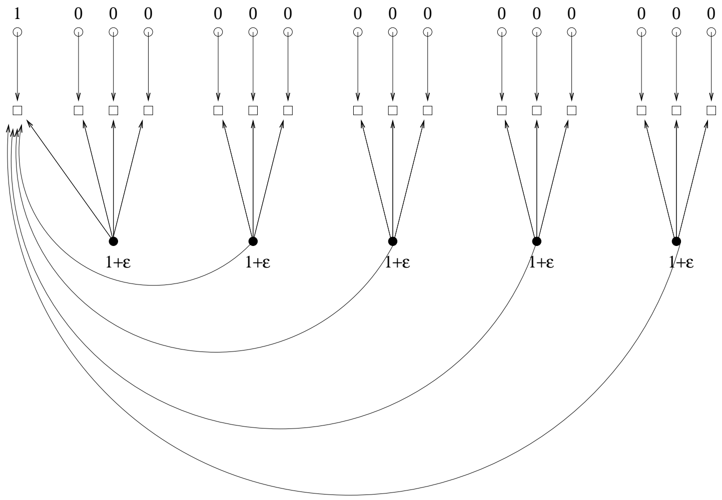

Fig. 3. An example where $p(G)$ is $\Omega(kq)$ higher than $p(OPT)$, for $k = 4$ and $q = 5$.

We can bound the payments of the approximation algorithm as follows:

Theorem 6. *Let $k \geq 2$ be the size of the largest set in a Set Cover instance, and $q \geq 2$ be an upper bound on the number of sets that contain an element. Then $p(G) \leq qk^2 \cdot p(OPT)$.*

Proof. We assume without loss of generality that $p(OPT)$ is finite, which implies that for every set S in the optimum cover, there is a cover without S.

We use arguments and notation from Theorem 3 repeatedly. Let $B_1, B_2, \ldots, B_m$ be the sets picked by GREEDY in this order. Let Q_i be the subset of B_i with the elements first covered by B_i (that is, not covered by $B_1, B_2, \ldots, B_{j-1}$), and note that for $1 \le i < j \le m$, we have $Q_i \cap Q_j = \emptyset$.

Consider a set B_i and we analyze $p(B_i)$. We increase $c(B_i)$ until GREEDY does not pick B_i. We reach the following situation: GREEDY picks $B_1, B_2, \ldots, B_{i-1}$, then it might pick a several sets H_j, and then it picks B_i. However, a further increase in $c(B_i)$ results in B_i not being picked. Let $\overline{c}$ be this new cost function which differs from c only for B_i, and note that $p(B_i) = \overline{c}(B_i)$. Define $\overline{Q}_i$ to be the elements first covered by B_i in the cover given by GREEDY when run with $\overline{c}$, and note that $\overline{Q}_i \subseteq Q_i$.

Define Z_i to be the set of elements not covered by $OPT \setminus \{B_i\}$ (so $Z_i = \emptyset$ if $B_i \notin OPT$). If $B_i \in OPT$, define M_i and N_i to be collection of sets such that $M_i \cup N_i$ is a minimum cost cover which does not use B_i and with M_i covering $Z_i \cap \overline{Q}_i$. For convenience define $M_i = \emptyset$ if $B_i \notin OPT$. Define U_i to be some collection of sets from OPT which cover $\overline{Q}_i \setminus Z_i$.

The arguments from Theorem 3 give:

$$p(G, B_i) \le \sum_{R_j \in M_i} c(R_j) + \sum_{R_j \in U_i} c(R_j). \tag{4}$$

Also, each $R_j \in OPT$ appears for at most k indexes i in sets U_i as the sets $\overline{Q}_i$ are disjoint and R_j has at most k elements. Note that, as in the VCG mechanism,

$$p(OPT, B_i) = c(M_i) + c(N_i) - c(OPT) - c(B_i), \tag{5}$$

where $c(M_i) = \sum_{R_j \in M_i} c(R_j)$ and $c(N_i) = \sum_{R_j \in N_i} c(R_j)$. Let V_i be the collection of sets of $OPT \setminus \{B_i\}$ which cover $\cup_{R_j \in M_i} R_j \setminus B_i$.

Consider the set of elements F which are not covered by any set of $V_i \cup \{B_i\}$. We have two covers for F: $OPT \setminus V_i \setminus \{B_i\}$ and N_i. The optimality of OPT implies that $c(N_i) \ge c(OPT) - c(V_i) - c(B_i)$. Plugging this into Equation 5 gives:

$$c(M_i) = p(OPT, B_i) - c(N_i) + c(OPT) + c(B_i) \le p(OPT, B_i) + 2c(B_i) + c(V_i). \tag{6}$$

We say that B_i charges S, for set $S \in OPT$, if $S \in V_i$. We count how many times can S be charged. For B_i to charge S, there must be a set R and an element $e \in \overline{Q}_i \cap R$ such that $R \cap S \neq \emptyset$. There are at most $k(q-1)$ such sets R, and not counting the elements of S, the cover at most $k(k-1)(q-1)$ elements. Noting that the sets $\overline{Q}_i$ are disjoint and therefore no two B_i can charge S through the same e, we obtain that S can be charged at most $k(k-1)(q-1)$ times.

Now Equations 4 and 6 give

$$p(G, B_i) \le c(U_i) + p(OPT, B_i) + 2c(B_i) + c(V_i) \tag{7}$$

. Summing over i and taking into account the number of charges, we obtain

$$p(G) \le kc(OPT) + p(OPT) + 2c(OPT) + k(k-1)(q-1)c(OPT). \tag{8}$$

As $c(OPT) \le p(OPT)$ and $k, q \ge 2$, we conclude that $p(G) \le qk^2 p(OPT)$. ■

5 Acknowledgments

We are grateful to Xiang-Yang Li and Weizhao Wang for providing us with an early draft of [16], thus inspiring this work. We also thank Bill Cunningham for discussions on matroids.

References

1. A. Archer, C. Papadimitriou, K. Talwar and E. Tardos, "An Approximate Truthful Mechansim for Combinatorial Auctions with Single Parameter Agents," *Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms* , pp. 205–214, 2003.
2. A. Archer and E. Tardos. "Truthful Mechanisms for One-Parameter Agents," *Proceedings of the 42nd IEEE symposium on Foundations of Computer Science*, pp. 482–491, 2001.
3. A. Archer and E. Tardos. "Frugal path mechanisms," *Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms* , pp. 991–999, 2002.
4. G. Baudis, C. Gropl, S. Hougardy, T. Nierhoff, and H. J. Promel. "Approximating minimum spanning sets in hypergraphs and polymatroids," *ICALP*, 2000.
5. P. Berman and V. Ramaiyer. "Improved Approximations for the Steiner Tree Problem," *J. Algorithms*, vol. 17, pp. 381-408, 1994.
6. S. Bikhchandani, S. de Vries, J. Schummer, and R. Vohra. "Linear programming and Vickrey auctions," *IMA Volumes in Mathematics and its Applications, Mathematics of the Internet: E-Auctions and Markets* vol. 127, pp. 75–116, 2001.
7. P.M. Camerini, G. Galbiati, and F. Maffioli. "Random pseudo-polynomial algorithms for exact matroid problems," *J. Algorithms*, vol. 13, pp. 258–273, 1992.
8. V. Chvatal. "A greedy heuristic for the set covering problem," *Mathematics of Operation Research*, vol. 4, pp. 233-235, 1979.
9. E. H. Clarke. "Multipart pricing of public goods," *Public Choice*, vol. 8, pp. 17–33, 1971.
10. W. J. Cook, W. H. Cunningham, W. R. Pulleyblank, and A. Schrijver. *Combinatorial Optimization.* Wiley-Interscience, 1998.
11. E. Elkind, A. Sahai, and K. Steiglitz. "Frugality in path auctions," *Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms* , pp. 701–709, 2004.
12. T. Groves. "Incentive in teams," *Econometrica*, vol. 41(4), pp. 617–631, 1973.
13. T. Helgason. "Aspects of the theory of hypermatroids," *Hypergraph Seminar: Ohio State University*, pp. 191–213, 1974.
14. S. Hougardy and H. J. Promel. "A 1.598 approximation algorithm for the Steiner problem in graphs," *Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms* , pp. 448–453, 1999.
15. M. Karpinski and A. Zelikovsky. "New Approximation Algorithms for the Steiner Tree Problems," *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 2(030), 1995.
16. X.-Y. Li and W. Wang. "Efficient Strategyproof Multicast in Selfish Networks," *International Workshop on Theoretical and Algorithmic Aspects of Sensor, Ad Hoc Wireless and Peer-to-Peer Networks*, 2004.
17. L. Lovasz. "Flats in matroids and geometric graphs," *Sixth British combinatorial conference*, pp. 45–86, 1977.
18. A. Mas-Colle, W. Whinston, and J. Green. *Microeconomic Theory* Oxford university press, 1995.
19. C. McDiamid. "Rado's theorem for polymatroids," *Math. Proc. Cambridge Philos. Soc.*, vol. 78, pp. 263–281, 1975.
20. N. Nisan and A. Ronen. "Algorithmic mechanism design," *Proceedings of the thirty-first annual ACM symposium on Theory of computing* , pp. 129–140, 1999.
21. C. Papadimitriou. "Algorithms, games, and the Internet," *Proceedings of the thirty-third annual ACM symposium on Theory of computing* , pp. 749–753, 2001.
22. H.J. Promel and A. Steger. "A new approximation algorithm for the Steiner tree problem with performance ratio 5/3," *J. of Algorithms*, vol. 36, pp. 89–101, 2000.
23. G. Robins and A. Zelikovsky. "Improved Steiner Tree Approximation in Graphs," *Proc. ACM-SIAM Symposium on Discrete Algorithms*, pp. 770–779, 2000.
24. A. Ronen. "Algorithms For Rational Agents," *SOFSEM*, 2000.
25. K. Talwar. "The price of truth: Frugality in truthful mechanisms," 20^{th} *Annual Symposium on Theoretical Aspects of Computer Science* , pp. 608–619, 2003.
26. W. Vickrey. "Counterspeculation, auctions and competitive sealed tenders," *Journal of Finance*, vol. 16, pp. 8–37, 1961.
27. L.A. Wolsey. "Analysis of the greedy algorithm for the submodular set covering problem," *Combinatorica*, vol. 2, pp. 385–392, 1982.
28. A. Zelikovsky. "An 11/6-Approximation Algorithm for the Network Steiner Tree Problem," *Algorithmica*, vol. 9, pp. 463–470, 1993.