The Budgeted Unique Coverage Problem and Color-Coding (Extended Abstract)

Neeldhara Misra[1], Venkatesh Raman[1], Saket Saurabh[2], and Somnath Sikdar[1]

[1] The Institute of Mathematical Sciences, Chennai, India.
{neeldhara|vraman|somnath}@imsc.res.in
[2] University of Bergen, Bergen, Norway.
saket.saurabh@ii.uib.no

Abstract. We show, by a non-trivial application of the color-coding method of Alon et al. [2], that BUDGETED UNIQUE COVERAGE (a variant of SET COVER) is fixed-parameter tractable, answering an open problem posed in [13]. We also give improved fixed-parameter tractable algorithms for two special cases of BUDGETED UNIQUE COVERAGE: UNIQUE COVERAGE (the unweighted version) and BUDGETED MAX CUT.

To derandomize our algorithms we use an interesting variation of k-perfect hash families known as (k, s)-hash families which were studied by Alon et al. [1] in the context of a class of codes called parent identifying codes [3]. In this setting, for every s-element subset S of the universe, and every k-element subset X of S, there exists a function that maps X injectively and maps the remaining elements of S into a different range.

We give several bounds on the size of (k, s)-hash families. We believe that our application of color-coding may be used for other problems and that this is the first application of (k, s)-hash families to a problem outside the domain of coding theory.

1 Introduction

The UNIQUE COVERAGE problem is a variant of SET COVER where, given a family of subsets of a finite universe, one is interested in finding a subfamily that maximizes the number of elements *uniquely* covered. This problem is motivated by a real-world application arising in wireless networks and has connections to several problems including MAX CUT and MAXIMUM COVERAGE [10].

Demaine et al. [5] introduced this problem and gave efficient approximation algorithms and inapproximability results. Moser et al. [13] studied the parameterized complexity of UNIQUE COVERAGE. They show that the problem is fixed-parameter tractable when parameterized by the number of elements to be uniquely covered. In particular, they left open the parameterized complexity of the more general version where elements have integral profits and sets have integral costs and one is interested in maximizing the total profit of elements uniquely covered by a minimum cost subfamily. In this paper, we show that (the standard parameterized version of) BUDGETED UNIQUE COVERAGE is fixed-parameter tractable. We also give improved algorithms for two special cases of BUDGETED UNIQUE COVERAGE: UNIQUE COVERAGE, the unweighted version of the problem and BUDGETED MAX CUT, a weighted variant of the well-known MAX CUT problem. See [7] and [9] for other related work on the UNIQUE COVERAGE problem.

In the BUDGETED UNIQUE COVERAGE problem, we are given a universe, where each element has a positive integral profit and a family of subsets of the universe, where each set has a positive integral cost. The question is whether there is a subfamily with total cost at most B that uniquely covers elements with total profit at least k. We show that this problem is fixed-parameter tractable with parameters k and B using the color-coding technique introduced by Alon et al. [2]. It is possible to derandomize the algorithm using standard s-perfect hash families where s is the maximum number of elements in a solution subfamily. However, we can use a variation of s-perfect families called (k, s)-hash families which were introduced in the context of a class of codes called parent identifying codes [3, 1]. To the best of our knowledge, we provide the first application of this class of hash families outside the domain of coding theory.

The rest of this paper is organized as follows. In Section 2, we apply color-coding to BUDGETED UNIQUE COVERAGE and show that it is fixed-parameter tractable. This section also contains the description of the hash families we use for derandomization. In Section 3 we consider two special cases of BUDGETED UNIQUE COVERAGE: UNIQUE COVERAGE and BUDGETED MAX CUT, and give better deterministic algorithms for these problems than the ones presented in [13]. We conclude with some open problems in Section 4. Complete proofs appear in a full version of this paper [12].

A parameterized problem is *fixed-parameter tractable (FPT)* if there is an algorithm that takes as input an instance (x, k) of the problem and correctly decides whether it is a YES or NO-instance in time $O(f(k) \cdot |x|^{O(1)})$, where f is some arbitrary function of the parameter k. When there is more than one parameter, k would represent an appropriate function (the sum or the maximum of them, for example) of the parameters. For further details and an introduction to parameterized complexity, we refer to [6, 8, 14]. For an integer n, by $[n]$ we denote the set $\{1, 2, \ldots n\}$. We let e denote the base of the natural logarithm (denoted by $\ln$) and $\log$ denote logarithms to base 2. We let $\mathbb{Q}$ denote the set of rationals, $\mathbb{Z}$ the set of integers, and for a number a, we use $\mathbb{Q}^{\geq a}$ to denote the set $\{x \in \mathbb{Q} : x \geq a\}$.

2 Budgeted Unique Coverage

An instance of UNIQUE COVERAGE consists of a family $\mathcal{F}$ of m subsets of a finite universe $\mathcal{U}$ of size n and a nonnegative integer k. The question is whether there exists a subfamily $\mathcal{F}' \subseteq \mathcal{F}$ that covers k elements uniquely. An element is said to be covered uniquely by $\mathcal{F}'$ if it appears in exactly one set of $\mathcal{F}'$. An instance of BUDGETED UNIQUE COVERAGE contains, in addition to $\mathcal{U}$ and $\mathcal{F}$, a cost function $c : \mathcal{F} \to \mathbb{Z}^+$, a profit function $p : \mathcal{U} \to \mathbb{Z}^+$ and nonnegative integers k and B. The question, in this case, is whether there exists a subfamily $\mathcal{F}' \subseteq \mathcal{F}$ of total cost at most B such that the total profit of elements uniquely covered by $\mathcal{F}'$ is at least k.

In [13] it was shown that BUDGETED UNIQUE COVERAGE with arbitrarily small positive (rational) costs and profits is not fixed-parameter tractable with parameters B and k, unless P = NP. Further, BUDGETED UNIQUE COVERAGE is $W[1]$-hard when parameterized by the budget B alone, even when cost and profit functions are integral. In this paper, we assume that both costs and profits assume positive integral values and that both B and k are parameters. Let $(\mathcal{U}, \mathcal{F}, c, p, B, k)$ be an instance of BUDGETED UNIQUE COVERAGE. We may assume that for all $S_i, S_j \in \mathcal{F}$, $i \neq j$, we have

- $S_i \neq S_j$;
- $c(S_i) \leq B$;
- $|S_i| \leq k - 1$.

For if $c(S_i) > B$ then S_i cannot be part of any solution and may be discarded; if $|S_i| \geq k$ then the given instance is trivially a YES-instance. We make these assumptions implicitly in the rest of the paper.

Demaine et al. [5] show that there exists an $\Omega(1/\log n)$-approximation algorithm for BUDGETED UNIQUE COVERAGE (Theorem 4.1). We use the same proof technique to show the following.

Lemma 1. *Let* $(\mathcal{U}, \mathcal{F}, c, p, B, k)$ *be an instance of* BUDGETED UNIQUE COVERAGE *and let* $c : \mathcal{F} \rightarrow \mathbb{Q}^{\geq 1}$ *and* $p : \mathcal{U} \rightarrow \mathbb{Q}^{\geq 1}$. *Then either*

1. *we can find in polynomial time a subfamily* $\mathcal{F}' \subseteq \mathcal{F}$ *with total cost at most* B *such that the total profit of elements uniquely covered by* $\mathcal{F}'$ *is at least* k*; or*
2. *for every subfamily* $\mathcal{H}$ *with total cost at most* B*, we have* $|\bigcup_{S \in \mathcal{H}} S| \leq 18k \log B$.

Proof. Appears in the full version [12]. □

The first step of our algorithm is to apply Step 1 of Lemma 1. From now on we assume that every subfamily of total cost at most B covers at most $18k \log B$ elements of the universe.

We now proceed to show that BUDGETED UNIQUE COVERAGE is FPT. We first show this for the case when the costs and profits are all one and then handle the more general case of integral costs and profits. Therefore let $(\mathcal{U}, \mathcal{F}, B, k)$ be an instance of BUDGETED UNIQUE COVERAGE with unit costs and profits. For this version of the problem, we have to decide whether there exists a subfamily $\mathcal{F}' \subseteq \mathcal{F}$ of size at most B that uniquely covers at least k elements. A subfamily $\mathcal{F}'$ of size at most B that uniquely covers at least k elements is called a *solution subfamily*.

To develop our color-coding algorithm, we use two sets of colors $\mathcal{C}_g$ and $\mathcal{C}_b$ with the understanding that the (good) colors from $\mathcal{C}_g$ are used for the elements that are uniquely covered and the (bad) colors from $\mathcal{C}_b$ are used for the remaining elements. In the present setting, $\mathcal{C}_g = \{1, \ldots, k\}$ and $\mathcal{C}_b = \{k + 1\}$.

Remark. For our algorithms, any subset of k colors can play the role of good colors. For ease of presentation, we fix a set of good and bad colors while describing our randomized algorithms. Our derandomized algorithms assume that any set of k colors may be good colors.

We now describe the notion of a *good configuration*. Given $h : \mathcal{U} \rightarrow \mathcal{C}_g \uplus \mathcal{C}_b$ and $\mathcal{F}' \subseteq \mathcal{F}$, define $h(\mathcal{F}') := \bigcup_{\{i \in S, S \in \mathcal{F}'\}} \{h(i)\}$ and $\mathcal{U}(\mathcal{F}') := \bigcup_{S \in \mathcal{F}'} S$.

Definition 1. *Given* $h : \mathcal{U} \rightarrow \mathcal{C}_g \uplus \mathcal{C}_b$ *and* $\mathcal{C}_g' \subseteq \mathcal{C}_g$*, we say that*

a. *$\mathcal{F}' \subseteq \mathcal{F}$ has a good configuration with respect to (wrt) h and $\mathcal{C}'_g$ if*
  1. *$h(\mathcal{F}') \cap \mathcal{C}_g = \mathcal{C}'_g$, and*
  2. *the elements of $\mathcal{U}(\mathcal{F}')$ that are assigned colors from $\mathcal{C}'_g$ have distinct colors and are uniquely covered by $\mathcal{F}'$.*

b. *$\mathcal{F}$ has a good configuration wrt h and $\mathcal{C}'_g$ if there exists a subfamily $\mathcal{F}'$ with a good configuration wrt h and $\mathcal{C}'_g$. Call $\mathcal{F}'$ a witness subfamily.*

A solution subfamily (for the unit costs and profits version) is a subfamily $\mathcal{F}' \subseteq \mathcal{F}$ with at most B sets and which uniquely covers at least k elements.

The next lemma shows that if h is chosen uniformly at random from the space of all functions $f : \mathcal{U} \to [k+1]$ and $(\mathcal{U}, \mathcal{F}, B, k)$ is a YES-instance of BUDGETED UNIQUE COVERAGE with unit costs and profits, then with high probability a solution subfamily $\mathcal{F}'$ has a good configuration wrt h and $\mathcal{C}_g$. Note that such a uniformly chosen h maps every element from $\mathcal{U}$ uniformly at random to an element in $[k+1]$.

Lemma 2. *Let $(\mathcal{U}, \mathcal{F}, B, k)$ be a YES-instance of* BUDGETED UNIQUE COVERAGE *with unit costs and profits and let $h : \mathcal{U} \to [k+1]$ be a function chosen uniformly at random. Then a solution subfamily $\mathcal{F}'$ has a good configuration wrt h and $\mathcal{C}_g$ with probability at least $2^{-k(18 \log B \log(k+1) - \log k + \log e)}$.*

Proof. Let $\mathcal{F}'$ be a solution subfamily with at most B sets that covers the elements $Q = \{i_1, \ldots, i_k\}$ uniquely. Then $p := |\mathcal{U}(\mathcal{F}')| \leq 18k \log B$, by Lemma 1. To complete the proof, we show that $\mathcal{F}'$ has a good configuration with respect to h and $\mathcal{C}_g$ with probability at least $2^{-k(18 \log B \log(k+1) - \log k + \log e)}$. For $\mathcal{F}'$ to have a good configuration, we must have $h(i) = k+1$ for all $i \in \mathcal{U}(\mathcal{F}') \setminus Q$ and $h(i_1), \ldots, h(i_k)$ a permutation of $1, \ldots, k$. The probability **Pr** that this happens is:

$$\begin{aligned} \mathbf{Pr} &= \tfrac{1}{(k+1)^{|\mathcal{U}(\mathcal{F}') \setminus Q|}} \times \tfrac{k!}{(k+1)^k} \quad \geq \left(\tfrac{k}{e}\right)^k \tfrac{1}{(k+1)^p} = e^{k \ln(k/e) - p \ln(k+1)} \\ &\geq e^{k \ln(k/e) - 18k \log B \ln(k+1)} \geq 2^{-k(18 \log B \log(k+1) - \log k + \log e)} \end{aligned}$$

□

Given a coloring h, how do we find out whether $\mathcal{F}$ has a good configuration wrt h and $\mathcal{C}_g$? We answer this next.

Finding a good configuration. Observe that if $\mathcal{F}$ has a good configuration wrt h and $\mathcal{C}_g$, then any witness subfamily $\mathcal{F}'$ covers at least k elements uniquely. To locate such a family of size at most B we use dynamic programming over subsets of $\mathcal{C}_g$. To this end, let W be a $2^k \times B$ array where we identify the rows of W with subsets of $\mathcal{C}_g$ and the columns with the size of a subfamily. For a fixed coloring function h, a subset $\mathcal{C}'_g \subseteq \mathcal{C}_g$ and $1 \leq i \leq B$, define $W[\mathcal{C}'_g][i]$ as follows:

$$W[\mathcal{C}'_g][i] = \begin{cases} 1, & \text{if there exists } \mathcal{F}' \subseteq \mathcal{F}, \text{ with } |\mathcal{F}'| \leq i, \text{ with a good con-} \\ & \text{figuration wrt } \mathcal{C}'_g \text{ and } h. \\ 0, & \text{otherwise.} \end{cases}$$

The entry corresponding to $W[\emptyset][i]$ is set to 1 for all $1 \leq i \leq B$, as a convention. We fill this array in increasing order of the sizes of subsets of $\mathcal{C}_g$. Let $\mathcal{T}$ be the family of all sets $S \in \mathcal{F}$ such that $h(S) \cap \mathcal{C}_g \subseteq \mathcal{C}'_g$. Let $g(S)$ denote the set of good colors used in S. Then $W[\mathcal{C}'_g][i] = \bigvee_{S \in \mathcal{T}} W[\mathcal{C}'_g \setminus g(S)][i-1]$.

The correctness of the algorithm is immediate. Clearly if $W[\mathcal{C}_g][B] = 1$, then a subfamily with at most B sets that uniquely covers at least k elements exists, and can be found out by simply storing the witness families $\mathcal{F}'$ for every entry in the table and backtracking. The time taken by the algorithm is $O(2^k Bmk)$, since the size of the array is $2^k B$ and each entry of the array can be filled in time $O(mk)$, where $m = |\mathcal{F}|$.

Lemma 3. *Let $(\mathcal{U}, \mathcal{F}, B, k)$ be an instance of* BUDGETED UNIQUE COVERAGE *with unit costs and profits and $h : \mathcal{U} \to \mathcal{C}$ a coloring function. Then we can find a subfamily $\mathcal{F}'$ of size at most B which has a good configuration wrt h and $\mathcal{C}_g$, if there exists one, in time $O(2^k Bmk)$.*

A randomized algorithm for BUDGETED UNIQUE COVERAGE with unit costs and profits is as follows.

1. Randomly choose a coloring function $h : \mathcal{U} \to \{1, \ldots, k+1\}$.
2. Apply Lemma 3 and check whether there exists a family $\mathcal{F}'$ of size at most B that is witness to a good configuration wrt h and $\mathcal{C}_g$. If such a family exists, return YES, else go to Step 1.

By Lemma 2, if the given instance is a YES-instance, the probability that a solution subfamily $\mathcal{F}'$ has a good configuration wrt a randomly chosen function $h : \mathcal{U} \to \mathcal{C}$ and $\mathcal{C}_g$ is at least $2^{-k(18 \log B \log(k+1) - \log k + \log e)}$. By Lemma 3, we can find such a subfamily in time $O(2^k Bmk)$.

Theorem 1. *Let $(\mathcal{U}, \mathcal{F}, B, k)$ be an instance of* BUDGETED UNIQUE COVERAGE *with unit costs and profits. There exists a randomized algorithm that finds a subfamily $\mathcal{F}'$ of size at most B covering at least k elements uniquely, if there exists one, in expected $O(2^{18k \log B \log(k+1)} \cdot Bmk)$ time.*

2.1 Improving the Run-time

It is clear that if a solution subfamily $\mathcal{F}'$ is to have a good configuration wrt a randomly chosen coloring function h and $\mathcal{C}_g$, then h must assign all the non-uniquely covered elements of $\mathcal{F}'$ the color in $\mathcal{C}_b$. Intuitively, if we increase the number of colors in $\mathcal{C}_b$, we increase the probability that a specific target subfamily has a good configuration wrt a randomly chosen coloring function. We formalize this intuition below. We need the following inequality whose proof we omit.

Lemma 4. *For all $t \geq 2k$, $\left(\frac{t-k}{t}\right)^t \geq (2e)^{-k}$.*

Lemma 5. *Let $(\mathcal{U}, \mathcal{F}, B, k)$ be a* YES*-instance of* BUDGETED UNIQUE COVERAGE *with unit costs and profits; let $\mathcal{C}_g = [k]$, $\mathcal{C}_b = \{k+1, \ldots, q\}$ and $\mathcal{C} = [q]$ so that $q \geq 2k$. If $h : \mathcal{U} \to \mathcal{C}$ is chosen uniformly at random then every solution subfamily $\mathcal{F}'$ with p elements of the universe has a good configuration wrt h and $\mathcal{C}_g$ with probability at least $e^{-k}\left(\frac{k}{q-k}\right)^k (2e)^{-\frac{kp}{q}}$.*

Proof. Let the set of elements uniquely covered by $\mathcal{F}'$ be $Q = \{i_1, \ldots, i_k\}$. For $\mathcal{F}'$ to have a good configuration, the function h must map every element of $\mathcal{U}(\mathcal{F}') \setminus Q$ to $\mathcal{C}_b$ and map Q to $\mathcal{C}_g$ injectively. Therefore the probability **Pr** that $\mathcal{F}'$ has a good configuration wrt $\mathcal{C}_g$ and a randomly chosen h is:

$$\mathbf{Pr} = \tfrac{(q-k)^{p-k}}{q^{p-k}} \times \tfrac{k!}{q^k} \geq \left(\tfrac{q-k}{q}\right)^p \left(\tfrac{1}{q-k}\right)^k k^k e^{-k}$$
$$\geq e^{-k}\left(\tfrac{k}{q-k}\right)^k \left(1 - \tfrac{k}{q}\right)^p \geq e^{-k}\left(\tfrac{k}{q-k}\right)^k (2e)^{-\frac{kp}{q}} \text{ (by Lemma 4).}$$

□

If $(\mathcal{U}, \mathcal{F}, B, k)$ is a YES-instance of BUDGETED UNIQUE COVERAGE with unit costs and profits then $p \leq 18k \log B$. Also observe that $B \geq 2$, for otherwise the given instance is a NO-instance. Setting $p = 18k \log B$ and $q = k + p$ in Lemma 5 we can show that a solution subfamily $\mathcal{F}'$ has a good configuration wrt a randomly chosen coloring function h and $\mathcal{C}_g$ with probability at least $2^{-8.2k - k \log\log B}$. Combining this with Lemma 3, we obtain:

Theorem 2. *Let $(\mathcal{U}, \mathcal{F}, B, k)$ be an instance of* BUDGETED UNIQUE COVERAGE *with unit costs and profits. Then we can find a subfamily $\mathcal{F}'$ of size at most B covering at least k elements uniquely, if there exists one, in $O(2^{9.2k + k \log\log B} \cdot Bmk)$ expected time.*

2.2 Derandomization

We now discuss how to derandomize the algorithms described in the last subsection. In general, randomized algorithms based on the color-coding method are derandomized using a suitable family of hash functions or "universal sets". We need a family of functions from $\mathcal{U}$ to $[t]$, where $t \geq k+1$, such that for all $S \subseteq \mathcal{U}$ of size $s = \lceil 18k \log B \rceil$ and all $X \subseteq S$ of size k, there exists a function h in the family which maps X injectively and the colors it assigns to the elements in $S \setminus X$ are different from the ones it assigns to those in X.

Such hash families are called (k, s)-hash families (with domain $[n]$ and range $[t]$) and they were introduced by Barg et al. [3] in the context of particular class of codes called parent identifying codes. At this point, we recall the definition of an (n, t, s)-*perfect hash family*. A family $\mathcal{H}$ of functions from $[n]$ to $[t]$ is called an (n, t, s)-perfect hash family if for every subset $X \subseteq [n]$ of size s, there is a function $h \in \mathcal{H}$ that maps X injectively. Note that an (n, t, s)-perfect hash family is a (k, s)-hash family with domain $[n]$ and range $[t]$, and a (k, s)-hash family with domain $[n]$ and range $[t]$ is an (n, t, k)-perfect hash family. Therefore (k, s)-hash families may be thought of as standing in between k-perfect and s-perfect hash families.

Our deterministic algorithm simply uses functions from these families $\mathcal{H}$ for coloring and is described below. Given an instance $(\mathcal{U}, \mathcal{F}, B, k)$ of BUDGETED UNIQUE COVERAGE with unit costs and profits, we let $n = |\mathcal{U}|$, $\mathcal{C} = [t]$, and s to be the closest integer to our estimate in Lemma 1, which is $O(k \log B)$.

Deterministic Algorithm
for each $h \in \mathcal{H}$ **do**
  for each subset $X \subseteq \mathcal{C}$ of size k **do**

1. Define $\mathcal{C}_g = X$ and $\mathcal{C}_b = \mathcal{C} \setminus X$;
2. Apply Lemma 3 and check whether there exists a subfamily $\mathcal{F}'$ of size at most B which has a good configuration wrt $\mathcal{C}_g$ and h;
3. **if** yes, then **return** the corresponding $\mathcal{F}'$;

return NO;

The correctness of the algorithm follows from the description—if a witness subfamily for the given $\mathcal{F}$ exists, at least one $h \in \mathcal{H}$ will color all the uniquely covered elements of the witness subfamily distinctly, thereby resulting in a good configuration. The running time of the algorithm is $O\left(|\mathcal{H}| \cdot \binom{t}{k} \cdot 2^k Bmk\right)$.

Alon et al. [1] provide explicit constructions of (k, s)-hash families when the range is $k+1$ and ks, respectively.

Theorem 3 (Alon et al. [1]). *There exists an absolute constant $c > 0$ such that for all $2 \leq k < s$ there is an explicit construction of a (k, s)-hash family $\mathcal{H}$ with domain $[n]$ and range $[k+1]$ of size at most $2^{ck \log s} \cdot \log_{k+1} n$. When the range is $[ks]$, there exists an explicit construction of a (k, s)-hash family of size $O(k^2 s^2 \log n)$.*

If $t = k+1$, then by the above theorem, the running time of our deterministic algorithm is $O(2^{O(k \log k + k \log\log B)} \cdot Bmk \cdot \log n)$; when $t = ks$, the running time works out to be $O(2^{O(k \log k + k \log\log B)} \cdot mk^5 \cdot B \log^2 B \cdot \log n)$.

Theorem 4. *Let $(\mathcal{U}, \mathcal{F}, B, k)$ be an instance of* BUDGETED UNIQUE COVERAGE *with unit costs and profits. Then we can find a subfamily $\mathcal{F}'$ of size at most B covering at least k elements uniquely, if there exists one, in time $O(2^{O(k \log k + k \log\log B)} \cdot Bmk \cdot \log n)$.*

We next give alternate running time bounds using standard s-perfect hash families for derandomizing our algorithm.

Theorem 5 ([2, 15, 4]). *There exist explicit constructions of (n, t, s)-perfect hash families of size $2^{O(s)} \log n$ when $t = s$, and of size $s^{O(1)} \log n$ when $t = s^2$. In fact, for the case $t = s$, an explicit construction of an s-perfect hash family of size $6.4^s \log^2 n$ in time $6.4^s n \log^2 n$ is known.*

For $t = s$, using the construction of s-perfect hash families by Chen et al. [4], we obtain a running time of $O(6.4^s \log^2 n \cdot \binom{s}{k} \cdot 2^k \cdot Bkm)$. Since $s = O(k \log B)$, this expression simplifies to of $O(2^{O(k \log B)} \cdot \log^2 n \cdot Bmk)$. For $t = s^2$, we can use a hash family of size $s^{O(1)} \log n$ [2], and the expression for the running time then works out to be $O(2^{O(k \log k + k \log\log B)} \cdot \log n \cdot Bmk)$. We thus have

Theorem 6. *Let $(\mathcal{U}, \mathcal{F}, B, k)$ be an instance of* BUDGETED UNIQUE COVERAGE *with unit costs and profits. Then we can find a subfamily $\mathcal{F}'$ of size at most B covering at least k elements uniquely, if there exists one, in time $O(f(k, B) \cdot \log^2 n \cdot Bmk)$, where $f(k, B) = \min\{2^{O(k \log B)}, 2^{O(k \log k + k \log\log B)}\}$.*

If we ignore constants, Theorem 6 gives a run-time which is at least as good as that in Theorem 4.

We now consider existential results concerning hash families. The following is known about (n, t, s)-hash families.

Theorem 7 ([11]). *For all positive integers $n \geq t \geq s \geq 2$, there exists an (n, t, s)-perfect hash family $\Delta(n, t, s)$ of size $e^{s^2/t} s \ln n$.*

Alon et al. [1] provided existential bounds for (k, s)-hash functions for the case when $t = k + 1$. If we assume that $s \geq 2k$, then their existential bound works out to $(2e)^s \cdot e^{k \ln s} \cdot s \ln n$. In the lemmas that follow, we provide existential bounds for an arbitrary range.

Lemma 6. *Let $k \leq s \leq n$ be positive integers and let $t \geq 2k$ be an integer. There exists a (k, s)-hash family $\mathcal{H}$ with domain $[n]$ and range $[t]$ of size $(2e)^{sk/t} \cdot s \log n$.*

Proof. Let $\mathcal{A} = \{h : [n] \to [t]\}$ be the set of all functions from $[n]$ to $[t]$. For $h \in \mathcal{A}$, $S \subseteq [n]$ of size s and $X \subseteq S$ of size k, define h to be (X, S)*-hashing* if h maps X injectively such that $h(X) \cap h(S \setminus X) = \emptyset$ and *not* (X, S)*-hashing* otherwise.

Fix $S \subseteq [n]$ of size s and $X \subseteq S$ of size k. The probability **Pr** that a function h picked uniformly at random from $\mathcal{A}$ is (X, S)-hashing, is given by:

$$\begin{aligned}
\mathbf{Pr} &= \frac{\binom{t}{k} k! (t-k)^{s-k}}{t^s} > \left(\frac{t}{k}\right)^k \cdot \left(\frac{k}{e}\right)^k \cdot \frac{1}{t^k} \cdot \left(\frac{t-k}{t}\right)^{s-k} \\
&= \frac{1}{e^k} \left(\frac{t-k}{t}\right)^{s-k} \geq \frac{1}{e^k} \cdot \left(\frac{1}{2e}\right)^{k(s-k)/t} \qquad \text{(By Lemma 4.)} \\
&\geq \left(\frac{1}{2e}\right)^{ks/t}.
\end{aligned}$$

The probability that the function h is not (X, S)-hashing is less than $1 - (2e)^{-ks/t}$. If we pick N functions uniformly at random from $\mathcal{A}$ then the probability that none of these functions is (X, S)-hashing is less than $(1 - (2e)^{-ks/t)})^N$. The probability that none of these N functions is (X, S)-hashing for some (S, X) pair is less than $\binom{n}{s}\binom{s}{k}(1 - (2e)^{-ks/t})^N$, which in turn is less than $n^s(1 - (2e)^{-ks/t})^N$. For this family of N functions to be (X, S)-hashing for every (S, X) pair, we would want $n^s(1 - (2e)^{-ks/t})^N$ to be at most one. A simple calculation yields that this will hold when $N \geq (2e)^{ks/t} s \log n$. □

Lemma 7. *Let $k \leq s \leq n$ be positive integers and let $t \geq k + 1$. There exists a (k, s)-hash family $\mathcal{H}$ with domain $[n]$ and range $[t]$ of size $2^{O(k \log(s/k))} \cdot s \log n$.*

Proof. Let $F = \Delta(n, m, s)$, the (n, m, s)-perfect hash family obtained from Theorem 7, where we set $m = \lceil s^2/(k \log(s/k)) \rceil$. Let G be a family of functions g_X from $[m]$ to $[t]$, indexed by k-element subsets X of $[m]$ as follows. The function g_X maps X in an one-one, onto fashion to $\{1, \ldots, k\}$ and maps an element of $[m] - X$ to an arbitrary element in $\{k + 1, \ldots, t\}$. Our required family T of functions from $[n]$ to $[t]$ is obtained by composing the families F and G. It is easy to see that T is an s-discriminating (n, t, k)-perfect hash family and has the claimed bound for its size. □

Note that Lemma 6 requires that $t \geq 2k$ and that for Lemma 7 we have no restriction on t. Also observe that had we an explicit construction of a (k, s)-hash family satisfying the bound in Lemma 6, then by setting $s = t = O(k \log B)$, we would have obtained a running time of $O(2^{O(k \log\log B)} \cdot s \log n)$ which is significantly better than that given in Theorem 6. We believe that this is motivation for studying explicit constructions of (k, s)-hash families for an arbitrary range.

2.3 Generalized Costs and Profits

The dynamic programming procedure described for the case of unit profits and costs can be scaled to handle the more general case when costs are positive integers and profits rational numbers ≥ 1 and vice versa. The modifications required are omitted from this extended abstract. We obtain the following result analogous to Theorem 6:

Theorem 8. *Let $(\mathcal{U}, \mathcal{F}, B, k)$ be an instance of* BUDGETED UNIQUE COVERAGE *with either integral costs and rational profits ≥ 1 or with rational costs ≥ 1 and integral profits. Then one can find a subfamily $\mathcal{F}'$ of total cost at most B that uniquely covers elements with total profit at least k, if there exists one, in time $O(f(k, B) \cdot Bmk \log^2 n)$, where $f(k, B) = \min\{2^{O(k \log B)}, 2^{O(k \log k + k \log\log B)}\}$.*

3 Faster Deterministic Algorithms for Special Cases

We now present faster deterministic algorithms than the ones presented in [13] for two special cases of BUDGETED UNIQUE COVERAGE: UNIQUE COVERAGE (the unbudgeted version) and BUDGETED MAX CUT.

3.1 Unique Coverage

An instance $(\mathcal{U}, \mathcal{F}, k)$ of UNIQUE COVERAGE can be viewed as an instance of BUDGETED UNIQUE COVERAGE where the costs and profits are all one and the budget $B = k$ as we do not need more than k sets to cover k elements uniquely. Using Theorem 6, we immediately obtain an algorithm with run-time $O(2^{O(k \log k)} \cdot |\mathcal{F}| \cdot k^2 \log n)$. In this subsection we present an algorithm for UNIQUE COVERAGE that runs in deterministic $O(2^{O(k \log\log k)} \cdot |\mathcal{F}| \cdot k + |\mathcal{F}|^2)$ time beating the $O(4^{k^2} \cdot |\mathcal{F}|)$ algorithm in [13]. We first need some lower bounds on the number of elements that can be uniquely covered in any instance of UNIQUE COVERAGE.

Define the *frequency* f_u of an element $u \in \mathcal{U}$ to be the number of sets in the family $\mathcal{F}$ that contain u. Let γ denote the maximum frequency, that is, $\gamma = \max_{u \in \mathcal{U}}\{f_u\}$.

Lemma 8. *There exists a subfamily $\mathcal{F}' \subseteq \mathcal{F}$ such that $\mathcal{F}'$ covers at least $n/(4e \log \gamma)$ elements uniquely. Furthermore, such a subfamily can be found in polynomial time.*

Proof. Similar to the proof of Lemma 1 and appears in the full version [12]. □

Lemma 9. *Let $M = \max_{S \in \mathcal{F}}\{|S|\}$. Then there exists a subfamily $\mathcal{F}'$ that covers at least $n/(8e \log M)$ elements uniquely. Furthermore, such a subfamily can be found in polynomial time.*

Proof. We begin by constructing a subfamily $\mathcal{F}'$ from $\mathcal{F}$ that is *minimal* in the sense that every set in $\mathcal{F}'$ covers at least one element in $\mathcal{U}$ uniquely. Such a subfamily is easily obtained, by going over every set in the family and checking if it has at least one element which is not contained in any other set. Let m' denote the size of the subfamily $\mathcal{F}'$. For the proof of the lemma we distinguish two cases based on m':

Case 1: $m' \geq n/2$. As the subfamily is minimal, by construction, we are immediately able to cover at least $n/2$ elements uniquely. Thus $\mathcal{F}'$ itself satisfies the claim of the lemma.

Case 2: $m' < n/2$. In this case, we first claim that $|\{u \in \mathcal{F}' : f(u) < M\}| \geq n/2$. If not, then there would be more than $n/2$ elements whose frequency is at least M, which implies that $\sum_{S \in \mathcal{F}'} |S| > Mn/2$. On the other hand, $\sum_{S \in \mathcal{F}'} |S|$ is clearly at most $M(n/2 - 1)$ (because there are strictly less than $n/2$ sets in the family and the size of any set in the family is bounded by M). The claim implies that there exists a set of at least $n/2$ elements whose frequency is less than M. Denote this set of elements by $\mathcal{V}$. Consider the family $\mathcal{F}''$ obtained from $\mathcal{F}'$ as follows: $\mathcal{F}'' = \{S \cap \mathcal{V} \mid S \in \mathcal{F}'\}$. Applying Lemma 8 to the instance $(\mathcal{V}, \mathcal{F}'')$, we obtain a subfamily $\mathcal{T}$ of $\mathcal{F}''$ that covers at least $n/(8e \log M)$ elements uniquely. The corresponding subfamily of $\mathcal{F}'$ will clearly cover the same set of elements uniquely in $\mathcal{U}$. This completes the proof of the lemma. □

Using these lower bounds on the number of elements that are uniquely covered, we can upper bound the size of a YES-instance of the UNIQUE COVERAGE problem as a function of the parameter k. Let $(\mathcal{U}, \mathcal{F}, k)$ be an instance of UNIQUE COVERAGE. If $k \leq n/8e\log(k-1)$, then there exists a subfamily that covers k elements uniquely. If not, we have $k > n/8e \log k$, which implies that $n < 8ek \log k$.

Lemma 10. *Let* $(\mathcal{U}, \mathcal{F}, k)$ *be an instance of* UNIQUE COVERAGE. *Then, in polynomial time, we can either find a subfamily covering at least k elements uniquely, or an equivalent instance where the size of the universe is* $O(k \log k)$.

An improved algorithm for UNIQUE COVERAGE first applies Lemma 10 and obtains an instance of UNIQUE COVERAGE, $(\mathcal{U}, \mathcal{F}, k)$, where $n = |\mathcal{U}| \leq O(k \log k)$. Now we examine all k-sized subsets X of the universe $\mathcal{U}$ and check whether there exists a subfamily that covers it uniquely. Let $X = \{u_{i_1}, u_{i_2}, \ldots, u_{i_k}\}$, and let h be a function that maps X injectively to $\{1, \ldots, k\}$ and each element in $\mathcal{U} \setminus X$ to the color $k+1$. Applying Lemma 3 to the instance $(\mathcal{U}, \mathcal{F}, B = k, k)$, with the coloring function h described above gives us an algorithm to find the desired $\mathcal{F}'$ in time $O(2^k k^2 m)$. Note that a factor of k can be avoided by directly applying dynamic programming over subsets of X. The size of $\mathcal{U}$ is upper bounded by $8ek \log k$ and hence the total number of subsets that need to be examined is at most $\binom{8ek \log k}{k}$, which is bounded above by $(8e \log k)^k \leq 2^{4.5k + k \log\log k}$. Combining this with the above discussion results in:

Theorem 9. *Given an instance* $(\mathcal{U}, \mathcal{F}, k)$ *of* UNIQUE COVERAGE, *one can find a subfamily that uniquely covers at least k elements, if there exists one, in time* $O(f(k) \cdot mk + m^2)$, *where* $f(k) = 2^{5.5k + k \log\log k}$.

3.2 Budgeted Max Cut

An instance of BUDGETED MAX CUT consists of an undirected graph $G = (V, E)$ on n vertices and m edges; a cost function $c : V \to \mathbb{Z}^+$; a profit function $p : E \to \mathbb{Z}^+$; and positive integers k and B. The question is whether there exists a cut $(T, V - T)$, $\emptyset \neq T \neq V$, such that the total cost of the vertices in T is at most B and the total profit of the edges crossing the cut is at least k. This problem can be modelled as an instance BUDGETED UNIQUE COVERAGE by taking $\mathcal{U} = E$ and $\mathcal{F} = \{S_v : v \in V\}$, where $S_v = \{e \in E : e \text{ is incident on } v\}$.

In [13], an algorithm with run-time $O((B^2 \cdot k \cdot 2^k)^{\min\{B,k\}} \cdot m^{O(1)})$ was described for BUDGETED MAX CUT. Here we develop an algorithm with run-time $O(2^{O(k)} \cdot Bmk \cdot \log^2 n)$. Given $S \subseteq V$, we let $c(S)$ denote the total cost of the elements of S. If $(S, V - S)$ is a cut in a graph G, then $p(S, V - S)$ is the total profit of edges across the cut. Define the profit $\hat{p}(v)$ of a vertex v to be the sum of the profits of all the edges incident on it.

Lemma 11. *If (G, B, k, c, p) is a* YES*-instance of* BUDGETED MAX CUT *then there exists a cut $(S, S - V)$ such that $c(S) \leq B$, $p(S, V - S) \geq k$, and $|\bigcup_{v \in S} S_v| \leq 4k$.*

Proof. Since we are given a YES-instance of the problem, there exists a cut (T, T') such that $c(T) \leq B$ and $p(T, T') \geq k$. Call a vertex v of T *redundant* if $p(T - v, T' \cup v) \geq k$. From (T, T'), obtain a cut (S, S') such that $S \subset T$ and S does not contain any redundant vertices. Observe that $c(S) \leq B$ and $p(S, S') \geq k$. For any $v \in S$, $\hat{p}(v) \leq k - 1$ and $p(S - v, S' \cup v) \leq k - 1$. Therefore $p(S, S') \leq 2k$. For $v \in S$, partition S_v as $I_v \uplus C_v$, where I_v is the set of edges incident on v that lie entirely in S and C_v are the edges that lie across the cut (S, S'). Clearly $p(I_v) \leq p(C_v)$, for otherwise, $p(S-v, S' \cup v) > p(S, S')$, a contradiction to the fact that S has no redundant vertices. Therefore $\sum_{v \in S} p(I_v) \leq \sum_{v \in S} p(C_v) \leq 2k$. This yields $\sum_{v \in S} \hat{p}(v) \leq 4k$. Since the profits are at least one, we have $|\bigcup_{v \in S} S_v| \leq 4k$. □

We use the deterministic algorithm outlined before Theorem 6 with $t = s = 4k$ and a $4k$-uniform perfect hash family by Chen et al. [4]. The running time then works out to $O(6.4^{4k} \log^2 n \cdot \binom{4k}{k} \cdot 2^k Bmk)$ which simplifies to $O(2^{13.8k} \cdot Bmk \cdot \log^2 n)$.

Theorem 10. *Let (G, B, k, c, p) be an instance of* BUDGETED MAX CUT*. Then we can find a cut (S, S') such that $c(S) \leq B$ and $p(S, S') \geq k$, if there exists one, in time $O(2^{13.8k} \cdot Bmk \cdot \log^2 n)$.*

4 Conclusions

In this paper we gave fixed-parameter tractable algorithms for BUDGETED UNIQUE COVERAGE and several of its variants. Our algorithms were based on an application of the well-known method of color-coding. Our randomized algorithms have good running times but the deterministic algorithms make use of either (k, s)-hash families or perfect hash families and this introduces large constants in the running times, a common enough phenomenon when derandomizing randomized algorithms using such function families [2]. Our use of (k, s)-hash families to derandomize our algorithms is perhaps the first application outside the domain of coding theory and it suggests the importance of such hash families. It will be interesting to explicitly construct (k, s)-hash families of size promised by Lemma 6 and explore other applications of our generalization of the color-coding technique.

Acknowledgements. We thank Srikanth Srinivasan for discussions about Lemma 6.

References

1. N. Alon, G. Cohen, M. Krivelevich, and S. Litsyn. Generalized hashing and parent-identifying codes. *Journal of Combinatorial Theory Series A*, 104(1):207–215, 2003.
2. N. Alon, R. Yuster, and U. Zwick. Color-coding. *J. ACM*, 42(4):844–856, 1995.
3. A. Barg, G. Cohen, S. Encheva, G. Kabatiansky, and G. Zémor. A hypergraph approach to the identifying parent property: The case of multiple parents. *SIAM Journal of Discrete Mathematics*, 14(3):423–431, 2001.
4. J. Chen, S. Lu, S.-H. Sze, and F. Zhang. Improved algorithms for path, matching, and packing problems. In *SODA*, pages 298–307, 2007.
5. E. D. Demaine, M. T. Hajiaghayi, U. Feige, and M. R. Salavatipour. Combination can be hard: approximability of the unique coverage problem. In *SODA*, pages 162–171, 2006.
6. R. G. Downey and M. R. Fellows. *Parameterized complexity*. Springer-Verlag, New York, 1999.
7. T. Erlebach and E. J. van Leeuwen. Approximating geometric coverage problems. In *SODA*, pages 1267–1276, 2008.
8. J. Flum and M. Grohe. *Parameterized Complexity Theory*. Texts in Theoretical Computer Science. An EATCS Series. Springer-Verlag, Berlin, 2006.
9. V. Guruswami and L. Trevisan. The complexity of making unique choices: Approximating 1-in-k sat. In *APPROX-RANDOM*, pages 99–110, 2005.
10. S. Khuller, A. Moss, and J. Naor. The budgeted maximum coverage problem. *Inf. Process. Lett.*, 70(1):39–45, 1999.
11. K. Mehlhorn. On the program size of perfect and universal hash functions. In *Proceedings of the 23th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 170–175. IEEE, 1982.
12. H. Moser, N. Misra, V. Raman, S. Saurabh, and S. Sikdar. The parameterized complexity of the Unique Coverage problem. Manuscript, 2009.
13. H. Moser, V. Raman, and S. Sikdar. The parameterized complexity of the unique coverage problem. In *ISAAC*, pages 621–631, 2007.
14. R. Niedermeier. *Invitation to fixed-parameter algorithms*, volume 31 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2006.
15. J. P. Schmidt and A. Siegel. The spatial complexity of oblivious k-probe hash functions. *SIAM J. Computing*, 19(5):775–786, 1990.