ISSN 1342-2812

Research Reports on Mathematical and Computing Sciences

Private Approximation of the Set Cover Problem

Masatoshi Yashiro and Keisuke Tanaka

December 2006, C–234

Department of
Mathematical and
Computing Sciences
Tokyo Institute of Technology

SERIES C: Computer Science

Private Approximation of the Set Cover Problem

Masatoshi Yashiro and Keisuke Tanaka

Dept. of Mathematical and Computing Sciences
Tokyo Institute of Technology
W8-55, 2-12-1 Ookayama Meguro-ku, Tokyo 152-8552, Japan
{yashiro2, keisuke}@is.titech.ac.jp

December 27, 2006

Abstract

Private approximation, introduced by Feigenbaum, Ishai, Malkin, Nissim, Strauss, and Wright, allows us to find approximate solutions with disclosing as little information as possible. In STOC 2006, Beimel, Carmi, Nissim, and Weinreb studied the private approximation for both the vertex cover and the max exact 3SAT problems. In this paper, we consider the set cover problem where the costs of all sets are polynomially bounded. We show that there exists neither a deterministic nor a randomized private approximation. We also consider the case that the frequencies of all elements are equal. We show that in this case there exist no deterministic private approximation.

Keywords: private approximation, set cover problem.

1 Introduction

Approximation algorithms can sometimes provide efficient solutions when no efficient exact computation is known. Approximation algorithms are widely studied, especially for NP-hard problems.

Feigenbaum, Ishai, Malkin, Nissim, Strauss, and Wright [3] introduced the notion of the private approximation of functions. Roughly speaking, an approximation function $\hat{g}$ is called private approximation with respect to the target function g, if $\hat{g}(x)$ reveals no more information about x than $g(x)$ does. More formally, there exists a probabilistic polynomial time simulator $\mathcal{M}$ such that the distribution of the simulation output $\mathcal{M}(g(x))$ is indistinguishable from $\hat{g}(x)$. They proposed a function (two-party protocol) which is the private approximation with respect to that for computing the hamming distance between two binary vectors. They also proposed the private approximations of several natural $\#P$-hard problems. After [3], several private approximaitons were proposed [4, 10, 7].

Halevi, Krauthbamer, Kushilevitz, and Nissim [5] discussed the private approximation of NP-hard problems. They proved that there exists no private approximation for computing the size of minimum vertex cover within approximation ratio $n^{1-\epsilon}$. Their proof used the *sliding-window reduction* that translates a SAT instance ϕ to an instance G of the vertex cover problem. If ϕ is satisfiable then G has the vertex cover of size z, otherwise any vertex cover for G is of size at least $z+1$. The definition of the private approximation in [5] is almost the same as that by Feigenbaum, Ishai, Malkin, Nissim, Strauss, and Wright [3].

Beimel, Nissim, Carmi, and Weinreb [1] studied the private approximation for both the vertex cover and the max exact 3SAT problems. In order to consider search problems, they proposed a definition of the private approximation which is different from that in [3]. In their definition, an algorithm $\mathcal{A}$ is a private approximation with respect to a privacy structure $\mathcal{R}$, which is an equivalent relation, if the outputs of executing $\mathcal{A}$ on two $\mathcal{R}$-equivalent inputs are computational indistinguishable. Under their definition, they showed that there exists neither a deterministic nor a randomized private approximation of the search problem for a minimum vertex cover within approximation ratio $n^{1-\epsilon}$.

In this paper we consider the private approximation of the set cover problem. In the previous paper [5, 1], only the vertex cover problem whose costs of all not fixed that vertices are fixed. In particular, we consider the set cover problem where the costs of all sets are polynomially bounded. We show that there exists neither a deterministic nor a randomized private approximation. We also consider the case that the frequencies of all elements are equal. We show that in this case there exist no deterministic private approximation.

2 The Set Cover Problem

In this section, we describe the set cover problem and the frequency.

Definition 2.1 (Set Cover Problem)**.** *Let U be a set of m elements, $\mathcal{S} = \{S_1, \ldots, S_n\}$ a collection of subsets of U, and c: $\mathcal{S} \to \mathbb{Q}^+$ a cost function. We say the set $C \in \{1, \ldots, n\}$ of indices is called a cover of U if the collection of S_i ($i \in C$) covers all elements in U, that is, $\bigcup_{i \in C} S_i = U$. Given $\langle U, \mathcal{S}, c \rangle$, the set cover problem is to find a minimum cost cover of U.*

Karp [9] showed that set cover problem is a NP-hard problem. One of the best and well known polynomial-time approximation algorithms is the greedy algorithm: at each step the subset that covers the largest number of remaining elements. Johnson [8] and Lovász [11] independently showed that the performance ratio of the greedy method is no worth than $\ln |S| + 1$. Halldórsson [6] and Duh and Fürer [2] has improved the upper bound to about $(\ln |S|)/2$.

Usually, the size of the instance $\langle U, \mathcal{S}, c \rangle$ of the set cover problem is considered as the number of the elements in U. In this paper, we consider the size of the instance of the set cover problem as the number of sets in $\mathcal{S}$. Therefore, the number of the elements in each set in $\mathcal{S}$ is restricted to polynomial of the number of the sets in $\mathcal{S}$.

In this paper, we consider "*polynomial-cost set cover problem*" where the cost of each set is polynomial in the problem size. Let "`Set Cover`" be a polynomial-cost set cover problem.

Definition 2.2 (Frequency)**.** *We define the frequency of an element to be the number of sets the element is in. A useful parameter is the frequency of the most frequent element. Let us denote this by f.*

We call the problem where all elements in U have the equivalent frequency as "*set cover problem with fixed frequency*" .

3 The Approximation and the Private Algorithm

First, we describe the definition of the approximation. The following definition of the approximation can be applied to *minimization* problems. The definition for maximization problems is similar.

Definition 3.1 (Approximation of the Search Problem)**.** *Let g be a function, $\mathcal{A}$ an algorithm for a search problem, and c a cost function. We say that $\mathcal{A}$ is an α-approximation of g if it runs in polynomial time and for all input x,*

$$\sum_{y \in \mathcal{A}(x)} c(y) \leq \alpha \sum_{y \in g(x)} c(y).$$

Next, we describe the definition of the private algorithm, following [1]. We describe the privacy structure which is necessary to define the private algorithm.

Definition 3.2 (Privacy Structure)**.** *A privacy structure $\mathcal{R} \subseteq \{0,1\}^* \times \{0,1\}^*$ is an equivalence relation on instances. For $\langle x, y \rangle \in \mathcal{R}$, we use the notation $x \equiv_{\mathcal{R}} y$.*

We only discuss on the privacy structures of the form $\mathcal{R} = \cup_{n \in \mathbb{N}} \mathcal{R}_n$, where $\mathcal{R}_n$ is an equivalence relation among the instances of size n, such as $\mathcal{S}$ with n sets.

We now define the private algorithm. We say that an algorithm $\mathcal{A}$ is private with respect to a privacy structure $\mathcal{R}$ if the results of executing $\mathcal{A}$ on two $\mathcal{R}$-equivalent inputs are computationally indistinguishable.

Definition 3.3 (Private Algorithm)**.** *Let $\mathcal{R}$ be a privacy structure. A probabilistic polynomial-time algorithm $\mathcal{A}$ is private with respect to $\mathcal{R}$ if for every polynomial-time algorithm $\mathcal{D}$ and for every positive polynomial $p(\cdot)$, there exists some $n_0 \in \mathbb{N}$ such that for every $x, y \in \{0,1\}^*$, $x \equiv_{\mathcal{R}} y$, and $|x| = |y| \geq n_0$,*

$$\left| \Pr\left[\mathcal{D}\left(\mathcal{A}\left(x\right), x, y\right) = 1\right] - \Pr\left[\mathcal{D}\left(\mathcal{A}\left(y\right), x, y\right) = 1\right] \right| \leq \frac{1}{p(|x|)}.$$

That is, when $x \equiv_{\mathcal{R}} y$, any algorithm $\mathcal{D}$ cannot distinguish if the input of $\mathcal{A}$ is x or y.

Next, in order to define the private approximation of the search problem, we define the privacy structure, following [1]. We can regard the decision and the search problems as follows by using the bivariate relation.

Definition 3.4. *A bivariate relation Q is polynomially bounded if there exists a constant c such that $|w| \leq |x|^c$ for every $\langle x, w \rangle \in Q$. The decision problem for Q is, given an input x, to decide if there exists an element w such that $\langle x, w \rangle \in Q$ or not. The search problem for Q is, given an input x, to find an element w such that $\langle x, w \rangle \in Q$ if such w exists.*

We now define the privacy structure of the search problem. We require that if two input values have the same set of answers of the search problem, the approximation algorithm should not be able to distinguish between them.

Definition 3.5 (Privacy Structure of the Search Problem)**.** *The privacy structure $\mathcal{R}_Q$ related to the relation Q is defined as follows: $x \equiv_{\mathcal{R}_Q} y$ if and only if*

- $|x| = |y|$,
- *$\langle x, w \rangle \in Q$ if and only if $\langle y, w \rangle \in Q$ for every w.*

That is, $x \equiv_{\mathcal{R}_Q} y$ if they have the same set of solutions.

Finally, we give two relations of the problems considered in this paper.

$\langle U_1,\ \mathcal{S}_1,\ c_1\rangle$		$\langle U_2,\ \mathcal{S}_2,\ c_2\rangle$	
$U_1 = \{e_1,\ e_2,\ e_3,\ e_4\}$		$U_2 = \{e_1,\ e_2,\ e_3,\ e_4\}$	
$\mathcal{S}_1 = \{S_1,\ S_2,\ S_3,\ S_4\}$		$\mathcal{S}_2 = \{S_1,\ S_2,\ S_3,\ S_4\}$	
$S_1 = \{e_1,\ e_2\}$	$c_1(S_1) = 4$	$S_1 = \{e_1,\ e_2,\ e_3\}$	$c_2(S_1) = 4$
$S_2 = \{e_2,\ e_3\}$	$c_1(S_2) = 2$	$S_2 = \{e_2,\ e_3\}$	$c_2(S_2) = 3$
$S_3 = \{e_3,\ e_4\}$	$c_1(S_3) = 1$	$S_3 = \{e_4\}$	$c_2(S_3) = 2$
$S_4 = \{e_1,\ e_3\}$	$c_1(S_4) = 2$	$S_4 = \{e_1,\ e_3\}$	$c_2(S_4) = 1$

Figure 1: $\langle U_1, \mathcal{S}_1, c_1\rangle \equiv_{\mathcal{R}_{\text{minSC}}} \langle U_2, \mathcal{S}_2, c_2\rangle$. Note that $|\mathcal{S}_1| = |\mathcal{S}_2|$. Both solutions of $\langle U_1, \mathcal{S}_1, c_1\rangle$ and $\langle U_2, \mathcal{S}_2, c_2\rangle$ are equivalent ($\{1, 3\}$ and $\{2, 3, 4\}$).

Definition 3.6. *Let* minSC be the minimum set cover relation for `Set Cover`, that is, $\langle\langle U, \mathcal{S}, c\rangle, C\rangle \in$ minSC if C is the minimum cost cover for $\langle U, \mathcal{S}, c\rangle$. In this case, the privacy structure $\mathcal{R}_{\text{minSC}}$ contains all pairs $(\langle U_1, \mathcal{S}_1, c_1\rangle,\ \langle U_2, \mathcal{S}_2, c_2\rangle)$ where every minimum cost cover $C \in \mathcal{S}$ for $\langle U_1, \mathcal{S}_1, c_1\rangle$ is that for $\langle U_2, \mathcal{S}_2, c_2\rangle$ and vice verse. Similarly, let $\langle U, \mathcal{S}, c\rangle$ be the minimum cost cover relation for `Set Cover` with fixed-frequency.

In Figure 1, we give an example for the relation minSC.

4 Private Approximation of Set Cover

In this section, we show that there exists no deterministic f^{ϵ}-private approximation algorithm of `Set Cover`.

4.1 Definitions

In this section, we describe some definitions.

First, we describe the definition of the private approximation of `Set Cover`.

Definition 4.1 (Private Approximation of the Set Cover Problem)**.** *An algorithm $\mathcal{A}$ is a private α-approximation algorithm for* minSC if:

- $\mathcal{A}$ is a α-approximation algorithm for minSC, and
- $\mathcal{A}$ is private with respect to $\mathcal{R}_{\text{minSC}}$.

In order to analyze the private approximation of the vertex cover problem, Beimel et el. [1] employed "critical vertices" and "relevant vertices". We also employ the notion of "critical" and "relevant" for `Set Cover`.

Definition 4.2 (Critical Set and Relevant Set)**.** *Let U be a set of m elements, $\mathcal{S} = \{S_1, S_2, \ldots, S_n\}$ a collection of sets and c a cost function of $\mathcal{S}$. We say that S_i is critical for $\langle U, \mathcal{S}, c\rangle$ if every minimum set cover of $\langle U, \mathcal{S}, c\rangle$ contains S_i. We say that S_i is relevant for $\langle U, \mathcal{S}, c\rangle$ if there exists at least one minimum set cover of $\langle U, \mathcal{S}, c\rangle$ that contains S_i.*

Next, we present the problem related to Definition 4.2.

Definition 4.3 (The Relevant Set / Non-Critical Set Problem)**.**

Input: a Set U, a collection $\mathcal{S} = \{S_1, S_2, \ldots, S_n\}$ of sets, and a cost function c.

Output: "S_i is relevant for $\langle U, \mathcal{S}, c\rangle$" or "$S_i$ is non-critical for $\langle U, \mathcal{S}, c\rangle$".

It is easy to see that relevant include critical. Therefore some sets are both relevant set and non-critical. If the set is relevant and non-critical, it makes no difference to output relevant or non-critical.

We next define two special set cover problems. When we construct the algorithm for the Relevant / Non-Critical Set problem in Section 4.2, they are helpful.

Definition 4.4 ($\langle U^2, \mathcal{S}^2, c^2\rangle$ and $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$)**.** *Let U be a set that contains m elements $e_1, \ldots, e_m$, $\mathcal{S} = \{S_1, \ldots, S_n\}$ a collection of subsets of U, c a cost function $\mathcal{S} \to \mathbb{Q}^+$, I a collection of empty sets. For $\langle U, \mathcal{S}, c\rangle$ and for any $S_u \in I$ and $S_t \in \mathcal{S}$, we define $\langle U^2, \mathcal{S}^2, c^2\rangle$ and $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$ as follows.*

The collection $\mathcal{S}^2$ of sets is defined as $\mathcal{S}^2 = \{S_1, \ldots, S_{2m}\} \cup I$ where $S_{i+n} := \{e_{k+m} \mid e_k \in S_i\}$. The set U^2 is defined as $U^2 = \{e_1, \ldots, e_{2n}\}$. The function c^2 is defined as $c^2(S_i) = c(S_i)$ (for $1 \leq i \leq n$), $c^2(S_i) = c(S_{i-l})$ (for $l+1 \leq i \leq 2n$), and $c^2(S_i) = 1$ (for $S_i \in I$).

The collection $\mathcal{S}_{(t,u)}$ of sets is defined as $\mathcal{S}_{(t,u)} = \{S_1, \ldots, S_n\} \cup I$ where $S_t = S_t \cup \{e^*, e^{**}\}$, $S_u = S_u \cup \{e^*\}$, and $S_{u+n} = S_{u+n} \cup \{e^{**}\}$ for some $e^*, e^{**} \notin U^2$. The set $U_{(t,u)}$ is defined as $U_{(t,u)} = U^2 \cup \{e^*, e^{**}\}$, and let $c_{(t,u)} = c^2$.

We give concrete examples $\mathcal{S}^2$ and $\mathcal{S}_{(t,u)}$ in Figure 2.

4.2 Proofs

In this section, we show that there exists no deterministic f^ϵ-private approximation algorithm for `Set Cover` with respect to $\mathcal{R}_{\text{minSC}}$ if $P \neq NP$.

Theorem 4.5. *Let $\epsilon > 0$ be a constant and f a frequency. If $P \neq NP$, then there is no deterministic private f^ϵ-approximation algorithm of the search problem of* minSC .

This proof is similar to that in [1]. The outline of the proof is as follows :

1. We construct a Relevant or Non-Critical for Set Cover algorithm from the private approximation algorithm $\mathcal{A}$ with respect to $\mathcal{R}_{\text{minSC}}$.
2. We construct a greedy algorithm that efficiently solves the NP-hard problem from the Relevant or Non-Critical for Set Cover algorithm.
3. If $P \neq NP$, this is a contradiction. Thus there is no private approximation algorithm $\mathcal{A}$ for `Set Cover` with respect to $\mathcal{R}_{\text{minSC}}$.

In Algorithm 1, we describe a greedy algorithm of `Set Cover` given an access to the algorithm which decides relevant or non-critical (we call this algorithm Relevant or Non-Critical for Set Cover). We will show that the algorithm Relevant or Non-Critical for Set Cover can be constructed by using oracle access to private approximation algorithms of `Set Cover` later on.

$$\langle U, \mathcal{S}, c \rangle$$

$U = \{e_1, e_2, e_3, e_4\}$
$\mathcal{S} = \{S_1, S_2, S_3, S_4\}$
$S_1 = \{e_1, e_2\}$ $\quad c(S_1) = 4$
$S_2 = \{e_2, e_3\}$ $\quad c(S_2) = 1$
$S_3 = \{e_3, e_4\}$ $\quad c(S_3) = 2$
$S_4 = \{e_1, e_3\}$ $\quad c(S_4) = 2$

$$\Downarrow |I| = 2$$

$$\langle U^2, \mathcal{S}^2, c^2 \rangle$$

$U^2 = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8\}$
$\mathcal{S}^2 = \{S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8\} \cup I$
$I = \{S_9, S_{10}\}$
$S_1 = \{e_1, e_2\}$ $\quad c^2(S_1) = 4$
$S_2 = \{e_2, e_3\}$ $\quad c^2(S_2) = 1$
$S_3 = \{e_3, e_4\}$ $\quad c^2(S_3) = 2$
$S_4 = \{e_1, e_3\}$ $\quad c^2(S_4) = 2$
$S_5 = \{e_5, e_6\}$ $\quad c^2(S_5) = 4$
$S_6 = \{e_6, e_7\}$ $\quad c^2(S_6) = 1$
$S_7 = \{e_7, e_8\}$ $\quad c^2(S_7) = 2$
$S_8 = \{e_5, e_7\}$ $\quad c^2(S_8) = 2$
$S_9 = \emptyset$ $\quad c^2(S_9) = 1$
$S_{10} = \emptyset$ $\quad c^2(S_{10}) = 1$

$$\langle U_{(9,2)}, \mathcal{S}_{(9,2)}, c_{(9,2)} \rangle$$

$U_{(9,2)} = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8, e^*, e^{**}\}$
$\mathcal{S}_{(9,2)} = \{S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8\} \cup I$
$I = \{S_9, S_{10}\}$
$S_1 = \{e_1, e_2\}$ $\quad c_{(9,2)}(S_1) = 4$
$S_2 = \{e_2, e_3, e^*\}$ $\quad c_{(9,2)}(S_2) = 1$
$S_3 = \{e_3, e_4\}$ $\quad c_{(9,2)}(S_3) = 2$
$S_4 = \{e_1, e_3\}$ $\quad c_{(9,2)}(S_4) = 2$
$S_5 = \{e_5, e_6\}$ $\quad c_{(9,2)}(S_5) = 4$
$S_6 = \{e_6, e_7, e^{**}\}$ $\quad c_{(9,2)}(S_6) = 1$
$S_7 = \{e_7, e_8\}$ $\quad c_{(9,2)}(S_7) = 2$
$S_8 = \{e_5, e_7\}$ $\quad c_{(9,2)}(S_8) = 2$
$S_9 = \{e^*, e^{**}\}$ $\quad c_{(9,2)}(S_9) = 1$
$S_{10} = \emptyset$ $\quad c_{(9,2)}(S_{10}) = 1$

Figure 2: The constructions of $\langle U^2, \mathcal{S}^2, c^2 \rangle$ and $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)} \rangle$ $(t = 9,\ u = 2)$ using $\langle U, \mathcal{S}, c \rangle$ and I with size 2.

Algorithm. 1 (Greedy Minimum Set Cover)

Input: a collection of sets $\mathcal{S} = \{S_1, \ldots, S_n\}$, a cost function $c : \mathcal{S} \to \mathbb{Q}^+$, and a set U of m elements.

1. Set $C_s = \emptyset$.
2. If $U = \emptyset$ return C_s.
3. Pick a set $S_i \in \mathcal{S}$ and execute the algorithm Relevant or Non-Critical for Set Cover on $\langle U, \mathcal{S}, c\rangle$ and S_i.
4. If the answer is "Relevant",
   (a) Delete all the elements included in S_i from both U and sets S_j in $\mathcal{S}$. (We define these as $U^{'}$ and $\mathcal{S}^{'}$ for the sake of latter explanations.)
   (b) $\mathcal{S} \leftarrow \mathcal{S} \setminus \{S_i\}$.
   (c) $C_s \leftarrow C_s \cup \{i\}$.
   (d) Go to STEP2.
5. If the answer is "Non-Critical",
   (a) $\mathcal{S} \leftarrow \mathcal{S} \setminus \{S_i\}$.
   (b) Go to STEP2.

The following claim shows the correctness of the greedy algorithm.

Claim 4.6. *If the algorithm* Relevant or Non-Critical for Set Cover *is polynomial and correct then the algorithm* Greedy Minimum Set Cover *is polynomial and correct.*

proof. The algorithm is trivially correct for $U = \emptyset$. Now suppose $U \neq \emptyset$. If S_i is relevant for $\langle U, \mathcal{S}, c\rangle$, there is a minimum cover C for $\langle U, \mathcal{S}, c\rangle$ that contains S_i. Let d be a sum total of the cost of the set included in C. Then the set $C \setminus \{S_i\}$ is a cover of cost $d - c(S_i)$ for $\langle U^{'}, \mathcal{S}^{'} \setminus \{S_i\}, c\rangle$. By the induction hypothesis, the set C_v is a minimum cover for $\langle U^{'}, \mathcal{S}^{'} \setminus \{S_i\}, c\rangle$. We claim that $C_v \cup \{S_i\}$ is a minimum cover for $\langle U, \mathcal{S}, c\rangle$. All of the elements in S_i are covered by S_i, and the rest of the elements are covered by C_v. Since C_v is a minimum cover of $\langle U^{'}, \mathcal{S}^{'} \setminus \{S_i\}, c\rangle$, the cost of the covers is at most $d - c(S_i)$. Therefore, the cost of $C_v \cup \{S_i\}$ is d, and it is a minimum cover of $\langle U, \mathcal{S}, c\rangle$. If S_i is not critical for $\langle U, \mathcal{S}, c\rangle$, there is a minimum cover $C^{'}$ for $\langle U, \mathcal{S}, c\rangle$ that does not contain S_i. Therefore $C^{'}$ is one of the minimum covers for $\langle U^{'}, \mathcal{S}^{'} \setminus \{S_i\}, c\rangle$. It is oblivious that if the algorithm Relevant or Non-Critical for Set Cover is polynomial, then the greedy algorithm is polynomial. □

We next construct the Relevant or Non-Critical for Set Cover algorithm from a private approximation algorithm for minSC. We adopt the idea of [1].

Claim 4.7. *Let $\mathcal{S} = \{S_1, S_2, \ldots, S_n\}$ be a collection of sets, c a cost function $\mathcal{S} \to \mathbb{Q}^+$, e^* an element such that $e^* \notin U$. We choose $i, j \in \{1, \ldots, n\}$ arbitrary $i \neq j$, and define $\mathcal{S}^* = \{S_1^*, \ldots, S_n^*\}$ where $S_i^* = S_i \cup \{e^*\}$, $S_j^* = S_j \cup \{e^*\}$, and $S_k^* = S_k$ for $k \neq i$, j. We also define $U^* = U \cup \{e^*\}$ and $c^*(S_i^*) = c(S_i^*)$.*

Then, If S_j is critical for $\langle U, \mathcal{S}, c\rangle$, then $\langle U, \mathcal{S}, c\rangle \equiv_{\mathcal{R}_{\mathrm{minSC}}} \langle U^*, \mathcal{S}^*, c^*\rangle$.

proof. First, we show that a minimum set cover of $\langle U, \mathcal{S}, c\rangle$ is that of $\langle U^*, \mathcal{S}^*, c^*\rangle$. Let C be a minimum set cover of $\langle U, \mathcal{S}, c\rangle$. Since C contains (the index) j, and therefore C covers $\langle U^*, \mathcal{S}^*, c^*\rangle$. Further, every cover of $\langle U^*, \mathcal{S}^*, c^*\rangle$ covers $\langle U, \mathcal{S}, c\rangle$. Thus, C is a minimum set cover of $\langle U^*, \mathcal{S}^*, c^*\rangle$. Next, we show that every minimum set cover of $\langle U^*, \mathcal{S}^*, c^*\rangle$ is the minimum set cover of $\langle U, \mathcal{S}, c\rangle$. Let C^* be the minimum cover of $\langle U^*, \mathcal{S}^*, c^*\rangle$ and d the cost of the minimum cover of $\langle U, \mathcal{S}, c\rangle$. The minimum set cover C for $\langle U, \mathcal{S}, c\rangle$ contains j since S_j is critical. Thus, C is the minimum set cover for $\langle U^*, \mathcal{S}^*, c^*\rangle$, and the cost of C^* is at most d which is the cost of C. On the other hand, since $U \subseteq U^*$, C^* is a set cover of $\langle U, \mathcal{S}, c\rangle$ and the cost is at least d. Therefore the cost of C^* is d and C^* is the minimum cover of $\langle U, \mathcal{S}, c\rangle$. □

We prove the two claims with respect to $\langle U^2, \mathcal{S}^2, c^2\rangle$ and $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$ defined in Definition 4.4. We use these claims for the proof of the correctness of Algorithm 2.

Claim 4.8. *If S_u is critical for $\langle U, \mathcal{S}, c\rangle$, then $\langle U^2, \mathcal{S}^2, c^2\rangle \equiv_{\mathcal{R}_{\mathrm{minSC}}} \langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$.*

proof. Since $\mathcal{S}^2$ is composed of two separate copies of $\mathcal{S}$ and S_u is critical for $\langle U, \mathcal{S}, c\rangle$, then S_u and S_{u+n} are critical for $\langle U^2, \mathcal{S}^2, c^2\rangle$ where $m = |U|$. Therefore, by Claim 4.7, $\langle U^2, \mathcal{S}^2, c^2\rangle \equiv_{\mathcal{R}_{\mathrm{minSC}}} \langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$. □

Claim 4.9. *If S_u is not relevant for $\langle U, \mathcal{S}, c\rangle$, then S_t is critical for $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$.*

proof. Assume that S_t is not critical for $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$. Then, there exists a minimum set cover $C_{(t,u)}$ which does not contain S_t. Since e^* is contained nothing but S_t and S_u, $C_{(t,u)}$ must contain S_u. Similarly, $C_{(t,u)}$ must contain S_{u+n} in order to cover e^{**}. Then there is a minimum cover $C_{(t,u)}$ of $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$ that contains S_u and S_{u+n}. Therefore it is enough that we show that if S_u is not relevant for $\langle U, \mathcal{S}, c\rangle$, the cover $C_{(t,u)}$ of $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$ is not optimal.
Let d be cost of the minimum cover of $\langle U, \mathcal{S}, c\rangle$. Here, $C_{(t,u)}$ contains S_u and S_{u+n}, while the minimum cover of $\langle U, \mathcal{S}, c\rangle$ does not contain S_u if S_u is not relevant. Thus, $\sum_{i \in C(t,\ u)} c_{(t,u)}(S_i) \geq 2(d+1) = 2d+2$. Let C be the minimum cover of $\langle U, \mathcal{S}, c\rangle$, and $C_{+n} = \{s + n \mid s \in C\}$. Then the cost of C and that of C_{+n} are d, and $C \cup C_{+n} \cup \{t\}$ is a cover of $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$ of cost $2d+1$. This is a contradiction to the minimality of $C_{(t,u)}$. Therefore, S_t is critical for $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$. □

Next, by using a private f^ϵ-approximation algorithm we describe the Relevant or Non-Critical for Set Cover algorithm in Algorithm 2.

Algorithm. 2 (Relevant or Non-Critical for Set Cover)

Input: $(\langle U, \mathcal{S}, c\rangle,\ S_u)$

1. Let I be a set of $2df^{\epsilon}+1$ empty sets. (where $d = \sum_i c(S_i)$)
2. Construct the collection of sets $\mathcal{S}^2$ from $\mathcal{S}$ and I.
3. Execute $\mathcal{A}$ on $\langle U^2, \mathcal{S}^2, c^2\rangle$ and get the output W^2 of $\mathcal{A}$.
4. Choose any set $S_t \in I \setminus W^2$.
5. Construct the collection of sets $\mathcal{S}_{(t,u)}$ from $\mathcal{S}$, I, S_t, and S_u.
6. Execute $\mathcal{A}$ on $\mathcal{S}_{(t,u)}$, and get the output $W_{(t,u)}$ of $\mathcal{A}$.
7. If $W^2 \neq W_{(t,u)}$, return "Non-Critical". Else return "Relevant".

We show the following claim.

Claim 4.10. *Let $\mathcal{A}$ be a deterministic private approximation algorithm for* minSC, *U a set of m elements, $\mathcal{S} = \{S_1, \ldots, S_n\}$ a collection of subsets of U, and c a cost function, and denote $\mathcal{A}(\langle U, \mathcal{S}, c\rangle)$ be a cover of $\langle U, \mathcal{S}, c\rangle$ that corresponding to indices outputted by $\mathcal{A}$. Then for any set $i \in \mathcal{S} \setminus W$, the set S_i is not critical for $\langle U, \mathcal{S}, c\rangle$.*

proof. Let $U^{'}$ be $U \cup \{e\}$ where $e \notin U$, and $\mathcal{S}^{'} = \{S_1, \ldots, S_n\}$ where $S_i = S_i \cup \{e\}$. Let $W^{'}$ be the output of $\mathcal{A}$ with $\langle U^{'}, \mathcal{S}^{'}, c^{'}\rangle$. Then $W^{'} \neq W$ since S_i is included in $W^{'}$.
If S_i is critical for $\langle U, \mathcal{S}, c\rangle$, then, by Claim 4.9, the minimum set cover of $\langle U, \mathcal{S}, c\rangle$ and $\langle U', S', c'\rangle$ are equal. However since $\mathcal{A}$ is private and deterministic, if $W^{'} = W$, the minimum set cover of $\mathcal{S}$ and $\mathcal{S}^{'}$ must be difficult. It contradicts. Therefore S_i is not critical. □

We must prove Algorithm 2 is correct and running time is polynomial. We prove the correctness by proving the following two claims.

Claim 4.11. *If $W^2 \neq W_{(t,u)}$, then S_u is not critical for $\langle U, \mathcal{S}, c\rangle$.*

proof. Assume S_u is critical for $\langle U, \mathcal{S}, c\rangle$. Since, by Claim 4.7, $\langle U^2, \mathcal{S}^2, c^2\rangle \equiv_{\mathcal{R}_{\text{minSC}}} \langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$, the minimum set cover of $\langle U^2, \mathcal{S}^2, c^2\rangle$ and $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$ are equal. Therefore since $\mathcal{A}$ is private, $W^2 = \mathcal{A}(\mathcal{S}^2) = \mathcal{A}(\mathcal{S}_{(t,u)}) = W_{(t,u)}$. Hence if $W^2 \neq W_{(t,u)}$, S_u is not critical for $\langle U, \mathcal{S}, c\rangle$. □

Claim 4.12. *If $W^2 = W_{(t,u)}$, then S_u is relevant.*

proof. As $W^2 = W_{(t,u)}$ and $S_t \notin W^2$, then $S_t \notin W_{(t,u)}$. Therefore, by Claim 4.10, S_t is not critical for $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$. Since S_t is not critical for $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$, S_u is relevant for $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$. □

Finally, we show that there is a set chosen in STEP 4 of Algorithm 2.

Claim 4.13. *Let $\epsilon \geq 0$. There is a set $S_i \in I$ such that $S_i \in I \setminus W^2$.*

proof. The cost of minimum set cover of $\langle U^2, \mathcal{S}^2, c^2\rangle$ is twice the cost of the minimum set cover of $\langle U, \mathcal{S}, c\rangle$, thus, it is at most $2d$. Since $\mathcal{A}$ is an f^{ϵ}-approximation algorithm of `Set Cover`, the cost of $\mathcal{A}(\langle U^2, \mathcal{S}^2, c^2\rangle)$ is at most $2df^{\epsilon}$. Consequently, there is at least one set $S_i \in I \setminus W^2$. □

Therefore we proved Theorem 4.5.

Remark 4.14. *Since d is polynomial in the size of problem, the sizes of $\mathcal{S}^2$ and $\mathcal{S}_{(t,u)}$ are polynomial in the size of problem. Therefore we prove Theorem 4.5. When d is exponentially large, the sizes of $\mathcal{S}^2$ and $\mathcal{S}_{(t,u)}$ are are exponentially large. Therefore in this case, we do not get an impossibility result by applying our strategy.*

5 Randomized Private Approximation of the Set Cover Problem

In this section, we show that there exists no randomized private f^{ϵ}-approximation algorithm of `Set Cover`. The outline of the proof is similar to that in [1]. We execute the approximation algorithm k times to decide whether the set is Relevant or Non-Critical. We prove several claims than correspond to those in Section 4.

We use the Algorithm 1, and we use Algorithm 3 as Relevant or Non-Critical for Set Cover in Algorithm 1.

Claim 5.1. *Let $\mathcal{A}$ be a randomized private f^{ϵ}-approximation algorithm of* `Set Cover` *. For every polynomial $p(\cdot)$, there exists some $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$ and every triple $\langle U, \mathcal{S}, c\rangle$, where $|\mathcal{S}| = n$, if $\Pr\left[\{\mathrm{t}, \mathrm{u}\} \cap \mathcal{A}(\langle \mathrm{U}, \mathcal{S}, \mathrm{c}\rangle) = \emptyset\right] \geq \frac{1}{p(n)}$, for some $S_t, S_u \in \mathcal{S}$, then S_t is not critical for $\langle U, \mathcal{S}, c\rangle$.*

proof. Let n be an integer, $p(\cdot)$ be a polynomial, $\langle U, \mathcal{S}, c\rangle$ be a triple where U be a set of l elements $\{e_1, \ldots, e_l\}$, $\mathcal{S}$ be a set of n set $\{S_1, \ldots, S_n\}$, c be a cost function, and $S_t, S_u \in \mathcal{S}$. Assume that $\Pr\left[\{t, u\} \cap \mathcal{A}(\langle U, \mathcal{S}, c\rangle) = \emptyset\right]$. As S_t and S_u are both not in $\mathcal{A}(\langle U, \mathcal{S}, c\rangle)$ with probability at least $1/p(|n|)$, and $\mathcal{A}(\langle U, \mathcal{S}, c\rangle)$ is cover of $\langle U, \mathcal{S}, c\rangle$. Let $\mathcal{S}^* = \{S_1, \ldots, S_n\}$ where $S_t = S_t \cup e^*$ and $S_u = S_u \cup e^*$, $U^* = U \cup e^*$, and $c^* = c$. In every execution of $\mathcal{A}$ on $\langle U^*, \mathcal{S}^*, c^*\rangle$, the cover $\mathcal{A}(\langle U, \mathcal{S}, c\rangle)$ must cover the element e^*. Therefore $\Pr\left[\{t, u\} \cap \mathcal{A}(\langle U, \mathcal{S}, c\rangle) = \emptyset\right] - \Pr\left[\{t, u\} \cup \mathcal{A}(\langle U^*, \mathcal{S}^*, c^*\rangle) = \emptyset\right] \geq \frac{1}{p(|n|)}$.
Let algorithm $\mathcal{D}$, whose input is two triples and the output C of $\mathcal{A}$ on one of them, be following:

- If the number of pair of the sets, whose index is the same but elements is different, is more than two or exactly one, always return 1.
- If the number of pair of the sets, whose index is the same but elements is different, is exactly two (let two sets be S_t and S_u),
  - If C contains at least one of S_t or S_u, return 1.
  - Otherwise return 0.

Now consider the execution of $\mathcal{D}$ where the triples are $\langle U, \mathcal{S}, c\rangle$ and $\langle U^*, \mathcal{S}^*, c^*\rangle$ which differ in exactly two sets. Since $\mathcal{A}$ is private, there exists some n_0 such that $\mathcal{D}$ cannot distinguish between equivalent triples with more than n_0 sets. Thus, if $\langle U, \mathcal{S}, c\rangle$ has $n \geq n_0$ sets and $\Pr\left[\{t, u\} \cap \mathcal{A}(\langle U, \mathcal{S}, c\rangle) = \emptyset\right]$, the triples $\langle U, \mathcal{S}, c\rangle$ and $\langle U^*, \mathcal{S}^*, c^*\rangle$ are not equivalent. This implies, by Claim 4.7, that S_u is not critical for $\langle U, \mathcal{S}, c\rangle$. □

Algorithm. 3 (Randomized Relevant or Non-Critical for Set Cover)

Input: $(\langle U, \mathcal{S}, c\rangle,\ S_u)$.

1. If $\mathcal{S}$ contains less than n_2 sets (where $n_2 = \max\{n_0,\ n_1\}$), then find if S_u is relevant for $\langle U, \mathcal{S}, c\rangle$ or non-critical for $\langle U, \mathcal{S}, c\rangle$ using exhaustive search.
2. Let I be a set $4df^{\epsilon} + 2$ sets. (where $d = \sum_i c(S_i)$)
3. Construct the family of sets $\langle U^2, \mathcal{S}^2, c^2\rangle$ from $\langle U, \mathcal{S}, c\rangle$ and I as in Definition 4.8.
4. Execute k times the algorithm $\mathcal{A}$ on $\langle U^2, \mathcal{S}^2, c^2\rangle$.
5. Choose a set $S_t \in I$ such that S_t appears at most $k/2$ times in $\mathcal{A}(\langle U^2, \mathcal{S}^2, c^2\rangle)$ in the k executions.
6. Construct the family of sets $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$ from $\langle U, \mathcal{S}, c\rangle$, I, S_t, and S_u as in Definition 4.8.
7. Execute k times Algorithm $\mathcal{A}$ on $\mathcal{S}_{(t,u)}$.
8. If $t \in \mathcal{A}(\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle)$ in at least $0.75k$ of the k executions, then return "Non-Critical" Else return "Relevant".

We show the following claims.

Claim 5.2. *There is a set $S_t \in I$ such that index t appears at most $k/2$ times in $\mathcal{A}(\langle U^2, \mathcal{S}^2, c^2\rangle)$ out of the k executions.*

proof. The cost of minimum set cover of $\langle U^2, \mathcal{S}^2, c^2\rangle$ is twice the cost of the minimum set cover of $\langle U, \mathcal{S}, c\rangle$, thus, it is at most $2d$. Since $\mathcal{A}$ is an f^{ϵ}-approximation algorithm for `Set Cover` , the cost of $\mathcal{A}(\langle U^2, \mathcal{S}^2, c^2\rangle)$ is at most $2d \cdot f^{\epsilon} \leq \frac{|I|}{2}$. Thus, in each execution, at least half of the sets in I are not in $\mathcal{A}(\langle U^2, \mathcal{S}^2, c^2\rangle)$. Consequently, there is at least one set $S_t \in I$ such that t is not in $\mathcal{A}(\langle U^2, \mathcal{S}^2, c^2\rangle)$ in at least $k/2$ of the executions of $\mathcal{A}$ on $\langle U^2, \mathcal{S}^2, c^2\rangle$. □

Claim 5.3. *There exists a constant n_1 such that if*

- $\langle U^2, \mathcal{S}^2, c^2\rangle$ *contains at least n_1 sets,*
- $\Pr\left[t \in \mathcal{A}(\langle U^2, \mathcal{S}^2, c^2\rangle)\right] < 0.55$*, and*
- $\Pr\left[t \in \mathcal{A}(\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle)\right] > 0.6$,

then S_u is not critical for $\langle U, \mathcal{S}, c\rangle$

proof. Let Algorithm $\mathcal{D}$,whose input is two triples and the output C of $\mathcal{A}$ on one of them, be following:

- If the sets of empty sets in both triples are equal, returns 1.
- Otherwise, choose an empty set in exactly one of the family of sets, and
  - if this set is in C, return 1.
  - otherwise, return 0.

Now consider the execution of $\mathcal{D}$ where the triples are $\langle U^2, \mathcal{S}^2, c^2\rangle$ and $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$. Since $\mathcal{A}$ is private, there exists some $n_1 \in \mathbb{N}$ such that $\mathcal{D}$ cannot distinguish between equivalent triples with more than n_1 sets. Thus, if $\langle U^2, \mathcal{S}^2, c^2\rangle$ and $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$ have more than n_1 sets and $\Pr\left[t \in \mathcal{A}(\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle)\right] - \Pr\left[t \in \mathcal{A}(\langle U^2, \mathcal{S}^2, c^2\rangle)\right] \geq 0.05$, the sets $\langle U^2, \mathcal{S}^2, c^2\rangle$ and $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$ are not equivalent. This implies, by Claim 4.8, that S_t is non-critical for $\langle U, \mathcal{S}, c\rangle$. □

Claim 5.4. *If* $\Pr[\{t,u\} \cap \mathcal{A}\langle U, \mathcal{S}, c\rangle) = \emptyset] \leq 0.8$ *then* S_u *is relevant for* $\langle U, \mathcal{S}, c\rangle$.

proof. Since $|I| = 4df^\epsilon + 2$, there must be some same $S_k \in I$ such that $\Pr\left[t, k \cap \mathcal{A}(\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle) = \emptyset\right]$ is non negligible. By Claim 5.1, the set S_t is not critical for $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$. Hence, by Claim 4.9, the set S_t is relevant for $\langle U, \mathcal{S}, c\rangle$. □

Claim 5.5. *Let* $k > \Omega(df^\epsilon)$. *Algorithm* Randomized Relevant or Non-Critical for Set Cover *returns the correct answer with probability* $1 - O(2^{-k})$

proof. Let n_0 and n_1 be the constants guaranteed in Claim 5.3 and Claim 5.4 respectively, and define $n_2 = \max\{n_0, n_1\}$. If $\mathcal{S}$ contains less than n_2 sets, then the correctness is oblivious.
Let $S_w \in I$. By Chernoff bound, if $\Pr\left[t \in \mathcal{A}(\langle U^2, \mathcal{S}^2, c^2\rangle)\right] > 0.55$, the probability that $w \in \mathcal{A}(\langle U^2, \mathcal{S}^2, c^2\rangle)$ in less than $0.5k$ of the k executions of algorithm $\mathcal{A}$ on $\langle U^2, \mathcal{S}^2, c^2\rangle$ is $O(2^{-k})$. Therefore, the probability that $S_w \in I$ is include in $\langle U^2, \mathcal{S}^2, c^2\rangle$ such that $\Pr\left[t \in \mathcal{A}(\langle U^2, \mathcal{S}^2, c^2\rangle)\right] > 0.55$ and $w \in \mathcal{A}(\langle U^2, \mathcal{S}^2, c^2\rangle)$ in less than $0.5k$ of the k executions of Algorithm $\mathcal{A}$ on $\langle U^2, \mathcal{S}^2, c^2\rangle$ is $|I|\, O(2^{-k}) = df^\epsilon O(2^{-k}) = O(2^{-k})$ (since $k > \Omega(df^\epsilon)$). Therefore the probability that the set S_t chosen in Randomized Relevant or Non-Critical for Set Cover satisfies $\Pr\left[t \in \mathcal{A}(\langle U^2, \mathcal{S}^2, c^2\rangle)\right] < 0.55$ is at least $1 - O(2^{-k})$. We assume $\Pr\left[t \in \mathcal{A}(\langle U^2, \mathcal{S}^2, c^2\rangle)\right] < 0.55$ and prove the remainder.
Let $p = \Pr\left[t \in \mathcal{A}(\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle)\right]$. If $0.6 < p < 0.8$, then by Claim 5.3 and Claim 5.4, set S_u is both relevant and non-critical for $\langle U, \mathcal{S}, c\rangle$, thus the algorithm cannot error in this case. By Claim 5.3, If $p > 0.8$ then set S_u is not critical. If $p > 0.8$, by Chernoff bound, the probability that $t \in \mathcal{A}(\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle)$ in less than $0.75k$ of the k executions is $O(2^{-k})$, thus the algorithm errors with probability at most $O(2^{-k})$. By Claim 5.4, if $p < 0.6$ then set S_u is not critical. If $p < 0.6$, by chernoff bound, the probability that $u \in \mathcal{A}(\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle)$ in more than $0.75k$ of the executions is $O(2^{-k})$, thus the algorithm errors with probability at most $O(2^{-k})$. □

From the above claims, we can prove the following main theorem.

Theorem 5.6. *Let* $\epsilon > 0$ *be a constant. If RP* $\neq$ *NP, then there is no randomized private* f^ϵ*-approximation algorithm of* `Set Cover`.

proof. By Claim 5.5 and Claim 4.6, the success probability of the randomized algorithm of the exact search problems for minSC is $(1 - O(2^{-k}))^n$.

if there is a randomized private f^ϵ-approximation algorithm of `Set Cover`, then there is a randomized algorithm of the exact search problems for minSC. This algorithm is transformed to the algorithm of decision problem for `Set Cover` (given $\langle U, \mathcal{S}, c\rangle$ and $x \in \mathbb{Q}^+$, decide whether there is a cover of cost at most x). Since this problem is NP-complete, it contradicts $RP \neq NP$. □

6 Private Approximation of the Set Cover Problem with the Fixed Frequency

When the frequency of all elements U are 2, the set cover problem is essentially the same as the vertex cover problem. Therefore, in this section, we consider the situation that the frequencies of

all $e \in U$ are equal.

We show that there exists no randomized private approximation algorithm of `Set Cover` with the fixed frequency. The strategy of the proof is similar to that in the previous section, however, the construction of the greedy algorithm depends on whether the cost is fixed or not.

First, we describe the greedy algorithm for the case that the cost is not fixed. In this case, it is easy to construct the greedy algorithm.

Algorithm. 4 (Greedy Minimum Set Cover with fixed frequency (cost is not fixed))

Input: a collection of sets $\mathcal{S} = \{S_1, \ldots, S_n\}$, a cost function $c : \mathcal{S} \to \mathbb{Q}^+$, and a set U of m elements.

1. Set $C_s = \emptyset$.
2. If $U = \emptyset$ return C_s.
3. Pick an element $e_i \in U$ and make a list of all sets that include e_i. We define this list as $L = \{L_1, \ldots L_m\}$.
4. Set $c' \leftarrow c$ and $j = 1$.
5. Execute Algorithm Relevant or Non-Critical for Set Cover with fixed frequency on $\langle U, \mathcal{S}, c\rangle$ and $S_{j'}$ (where $S_{j'} = L_j$).
6. If the answer is "Relevant"
   (a) Delete all the elements included in $S_{j'}$ from both U and sets S_i in $\mathcal{S}$.
   (b) $\mathcal{S} \leftarrow \mathcal{S} \setminus \{S_{j'}\}$.
   (c) $c \leftarrow c'$.
   (d) $C_s \leftarrow C_s \cup \{j'\}$.
   (e) Go to STEP 2.
7. If the answer is "Non-Critical"
   (a) $c(S_{j'}) \leftarrow c(S_{j'}) + 1$.
   (b) $j \leftarrow j + 1$ and go to STEP 5.

We can show that this algorithm is polynomial and correct.

Claim 6.1. *If the algorithm* Relevant or Non-Critical for Set Cover with fixed frequency *is polynomial and correct and the cost of each set is not fixed then the algorithm* Greedy Minimum Set Cover with fixed frequency *is polynomial and correct.*

proof. This proof is the same proof of Claim 4.6 except when L_j is not critical. Thus we consider when L_j is not critical.

One is when L_i is not critical and not relevant(i.e. No optimal solution contain L_j) and the other is when L_j is not critical and relevant(i.e. At least one optimal solution contain L_j, but at least one optimal solution do not contain L_j). In the former, we can skip to L_j since every optimal solution do not contain S_i, and go to next iterate.

However in the latter, this algorithm is not guaranteed to stop if we skip to L_i. Since when

there are possibility that every sets that include e_i are non-critical, every sets are skipped, the algorithm can not output the solution.

In STEP 6, the cost of $S_{j'} = L_j$ is increased 1. This ensure that the all solutions that contain $S_{j'}$ become no optimal solution. (since $S_{j'}$ is not critical, there is at least one optimal solution without $S_{j'}$). In iteration, at least one in $L_1, \ldots, L_m$ become the critical set. Therefore this algorithm stop. Form Claim 4.6, Algorithm 4 is polynomial and correct if Algorithm Relevant or Non-Critical for Set Cover with fixed frequency is polynomial and correct . □

In Algorithm 4, if the cost is fixed, we can not execute STEP 6-a. Therefore we transform from Algorithm 4 to Algorithm 5.

Algorithm. 5 (Greedy Minimum Set Cover with fixed frequency (cost is fixed))

Input: a collection of sets $\mathcal{S} = \{S_1, \ldots, S_n\}$, a cost function $c : \mathcal{S} \to \mathbb{Q}^+$, and a set U of m elements.

1. Set $C_s = \emptyset$.
2. If $U = \emptyset$ return C_s.
3. Pick an element $e_i \in U$ and make a list of all sets that include e_i. We define this list as $L = \{L_1, \ldots, L_l\}$.
4. Set $\mathcal{S}' \leftarrow \mathcal{S}$, and $j = 1$.
5. If $j \leq l$,
   (a) Execute Algorithm Relevant or Non-Critical for Set Cover with fixed frequency on $\langle U, \mathcal{S}, c\rangle$ and $S_{j'}$ (where $S_{j'} = L_j$).
   (b) If the answer is "Relevant"
      i. Delete all the elements included in $S_{j'}$ from both U and all sets S_i in $\mathcal{S}$.
      ii. $\mathcal{S} \leftarrow \mathcal{S}' \setminus \{S_{j'}\}$.
      iii. $C_s \leftarrow C_s \cup \{j'\}$.
      iv. Go to STEP 2.
   (c) If the answer is "Non-Critical"
      i. If $|S_i| = 1$
         A. $j \leftarrow j + 1$ and go to STEP 5.
      ii. Else
         A. Divide S_i into $|S_i|$ $(= k)$ sets $S_{i_1}, \ldots, S_{i_k}$ (such that $|S_{i_\alpha}| = 1$ and if $\alpha \neq \beta$, $S_{i_\alpha} \cap S_{i_\beta} = \emptyset$).
         B. $\mathcal{S} \leftarrow (\mathcal{S} \setminus S_i) \cup S_{i_1} \cup \ldots \cup S_{i_k}$.
         C. $j \leftarrow j + 1$ and go to STEP 5.
6. If $j = l$,
   (a) $C_s \leftarrow C_s \cup h$ where $S_h = L_1$.
   (b) $\mathcal{S} \leftarrow \mathcal{S}' \setminus \{S_h\}$.
   (c) Go to STEP 2.

We show that this algorithm is polynomial and correctness.

Claim 6.2. *If Algorithm* Relevant or Non-Critical for Set Cover with fixed frequency *is polynomial and correct and the cost of each set is fixed, then Algorithm* Greedy Minimum Set Cover with fixed frequency *is polynomial and correct.*

proof. The proof correctness is similar to that of Claim 6.1. Along with Claim 6.1, we consider that L_j is not critical. Intuitively, by the set is divided, the set put off the optimal solution.

In the case of there exists only one element e such that $e \in S_j$ and $e \notin S$ $(S \neq S_j)$, there do

not exits critical set after STEP 5c-ii. However the set where include e is critical since other sets cover all elements except e. Thus by STEP 6a, we chose a set that include e.

Next, we consider that the case of there exists at least two elements e such that $e \in S_j$ and $e \notin S$ $(S \neq S_j)$. The cost of the cover that include S_j is increase since S_j is divided STEP 5c - ii (S_j need to cover at least two elements). Therefore similar to Claim 6.1, at least a set become critical set. Thus the correctness is proved.

This algorithm execute $O(n + mf)$ steps. However m and f are polynomial in n. Therefore the run time of this algorithm is polynomial in n. Thus Algorithm Greedy Minimum Set Cover with fixed frequency is polynomial. □

Next, we consider the algorithm Relevant or Non-Critical for Set Cover with fixed frequency. This is the Algorithm 2 where $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$ is replaced with $\langle U_{(t,\ u\ |\ f)}, \mathcal{S}_{(t,\ u\ |\ f)}, c_{(t,\ u\ |\ f)}\rangle$. We now define $\langle U_{(t,\ u\ |\ f)}, \mathcal{S}_{(t,\ u\ |\ f)}, c_{(t,\ u\ |\ f)}\rangle$.

Definition 6.3 ($\langle U_{(t,\ u\ |\ f)}, \mathcal{S}_{(t,\ u\ |\ f)}, c_{(t,\ u\ |\ f)}\rangle$)**.** *$\langle U^2, \mathcal{S}^2, c^2\rangle$ and I are the same as those defined by Definition 4.9. We choose the $2(f-1)$ elements from I, and which we denote $S_{k_1}, \ldots S_{k_{2(f-1)}}$. The collection $\mathcal{S}_{(t,\ u\ |\ f)}$ of sets is defined as $\mathcal{S}_{(t,\ u\ |\ f)} = \{S_1, \ldots, S_n\} \cup I$ where $S_t = \{e^*, e^{**}\}$, $S_u = S_u \cup \{e^*\}$, $S_{u+n} = S_{u+n} \cup \{e^{**}\}$, $S_{k_i} = \{e^*\}$ $(1 \leq i \leq f-1)$, and $S_{k_i} = \{e^{**}\}$ $(f \leq i \leq 2(f-1))$ such that $e^*, e^{**} \notin U^2$. The set $U_{(t,\ u\ |\ f)}$ defined as $U_{(t,\ u\ |\ f)} = U^2 \cup \{e^*, e^{**}\}$, and $c_{(t,\ u\ |\ f)} = c^2$.*

We can easily see $\langle U^2, \mathcal{S}^2, c^2\rangle$ and $\langle U_{(t,\ u\ |\ f)}, \mathcal{S}_{(t,\ u\ |\ f)}, c_{(t,\ u\ |\ f)}\rangle$ are `Set Cover` with fixed frequency if $\langle U, \mathcal{S}, c\rangle$ is `Set Cover` with fixed frequency.

Finally, we can prove Claims 6.4, 6.5, 6.6, and 6.7 in a similar way as those for the proof of Claims 4.8, 4.9, 4.11, and 4.12, respectively.

The following two claims are used in the proofs of Claim 6.6 and Claim 6.7.

Claim 6.4. *If S_u is critical for $\langle U, \mathcal{S}, c\rangle$, then $\langle U, \mathcal{S}, c\rangle \equiv_{\mathcal{R}_{\text{minSCfixed}}} \langle U_{(t,\ u\ |\ f)}, \mathcal{S}_{(t,\ u\ |\ f)}, c_{(t,\ u\ |\ f)}\rangle$.*

Claim 6.5. *If S_u is not relevant for $\langle U, \mathcal{S}, c\rangle$, then S_i is critical for $\langle U_{(t,\ u\ |\ f)}, \mathcal{S}_{(t,\ u\ |\ f)}, c_{(t,\ u\ |\ f)}\rangle$.*

The following two claims guaranteed the correctness of Algorithm 5.

Claim 6.6. *If $W^2 \neq W_{(t,\ u\ |\ f)}$, then S_u is not critical for $\langle U, \mathcal{S}, c\rangle$.*

Claim 6.7. *If $W^2 = W_{(t,\ u\ |\ f)}$, then S_u is relevant.*

We can prove the following theorem from the above claims and Claim 4.13.

Theorem 6.8. *Let $\epsilon > 0$ be a constant and f a frequency. If $P \neq NP$, then there is no deterministic private f^ϵ-approximation algorithm of the search problem for* minSCfixed.

7 Concluding Remarks

In this paper, we have considered the set cover problem where the costs of all sets are polynomially bounded. We have shown that there exists neither a deterministic nor a randomized private approximation. We have also considered the case that the frequencies of all elements are equal. We have shown that in this case there exist no deterministic private approximation.

In this paper, we have proved only when the size of the problem is defined as the number of the sets. It might be interesting to consider the problem where the size of the problem is defined as the number of elements. It might be also interesting to consider whether NP-hard problems other than the set cover problem have the private approximation algorithms or not.

Halevi et al. [5] discussed the leakage of information about the approximation algorithms for the minimum set cover problem. Beimel et al. [1] also discussed that for the vertex cover and the exact 3SAT problems. It might be interesting to consider the leakage of information about the approximation algorithms for the minimum set cover problem.

References

[1] Beimel, A., Carmi, P., Nissim, K., and Weinreb, E. Private Approximation of Search Problems. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC 2006)* (Seattle, WA, USA,, May 2006), ACM Press, pp. 119–128.

[2] chii Duh, R., and Fürer, M. Approximation of -Set Cover by Semi-Local Optimization. In *Proceedings on 29th Annual ACM Symposium on Theory of Computing (STOC '97)* (El Paso, Texas, USA, May 1997), ACM Press, pp. 256–264.

[3] Feigenbaum, J., Ishai, Y., Malkin, T., Nissim, K., Strauss, M., and Wright, R. N. Secure Multiparty Computation of Approximations. In *28th International Colloquium, ICALP 2001* (Crete, Greece, July 2001), F. Orejas, P. G. Spirakis, and J. van Leeuwen, Eds., vol. 2076, Springer-Verlag, pp. 927–938.

[4] Freedman, M. J., Nissim, K., and Pinkas, B. Efficient Private Matching and Set Intersection. In *Advances in Cryptology – EUROCRYPT 2004* (Interlaken, Switzerland, May 2004), C. Cachin and J. Camenisch, Eds., vol. 3027 of *Lecture Notes in Computer Science*, Springer-Verlag.

[5] Halevi, S., Krauthgamer, R., Kushilevitz, E., and Nissim, K. Private Approximation of NP-hard Functions. In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing (STOC 2001)* (Heraklion, Crete, Greece, July 2001), ACM Press, pp. 550–559.

[6] Halldórsson, M. M. Approximating Discrete Collections via Local Improvements. In *Proceedings of the 6th Annual ACM-SIAM Symposium on Discrete Algorithms – SODA '95* (San Francisco, California, January 1995), Lecture Notes in Computer Science, ACM Press, pp. 160–169.

[7] Indyk, P., and Woodruff, D. Polylogarithmic Private Approximations and Efficient Matching. In *Third Theory of Cryptography Conference – TCC 2006* (New York, NY, USA, March 2006), T. R. Shai Halevi, Ed., vol. 3876 of *Lecture Notes in Computer Science*, Springer-Verlag.

[8] Johnson, D. S. Approximation Algorithms for Combinatorial Problems. *Journal of Computer and System Sciences 9*, 3 (December 1974), 256–278.

[9] Karp, R. Reducibility among Combinatorial Problems. In *Complexity of Computer Computations*, R. Miller and J. Thatcher, Eds. Plenum Press, 1972, pp. 85–103.

[10] Kiltz, E., Leander, G., and Malone-Lee, J. Secure Computation of the Mean and Related Statistics. In *Third Theory of Cryptography Conference – TCC 2005* (Cambridge, MA, USA, February 2005), J. Kilian, Ed., vol. 3378 of *Lecture Notes in Computer Science*, Springer-Verlag.

[11] Lovász, L. On the Ratio of Optimal Integral and Fractional Covers. *Discrete Math. 13*, 4 (1975), 383–390.