A Note on the Density of the Multiple Subset Sum Problems

Yanbin Pan and Feng Zhang
Key Laboratory of Mathematics Mechanization,
Academy of Mathematics and Systems Science,
Chinese Academy of Sciences, Beijing, China, 100190
{panyanbin,zhangfeng}@amss.ac.cn

Abstract

It is well known that the general subset sum problem is NP-complete. However, almost all subset sum problems with density less than $0.9408\ldots$ can be solved in polynomial time with an oracle that can find the shortest vector in a special lattice. In this paper, we give a similar result for the multiple subset sum problem which has k subset sum problems with the same solution. A modified lattice is involved to make the analysis much simpler than before. In addition, some extended versions of the multiple subset sum problem are also considered.

Keywords. Density, Lattice, Multiple Subset Sum Problem

1 Introduction

The subset sum problem refers to the question to find variables $(x_1, x_2, \cdots, x_n) \in \{0,1\}^n$, given positive integers $a_1, a_2, \cdots, a_n$ and s, where s is the sum of some subset of the a_i's, such that

$$\sum_{i=1}^{n} x_i a_i = s.$$

The problem is well known to be NP-complete [2] and has many applications in cryptography, such as the Merkle-Hellman cryptosystem [6].

The density of these a_i's is defined by

$$d = \frac{n}{\log_2(\max_i a_i)}.$$

In terms of public-key cryptosystems, d is an approximate measure of the information rate at which bits are transmitted, namely

$$d \approx \frac{\text{number of bits in plaintext message}}{\text{average number of bits in ciphertext message}}.$$

It was shown by Lagarias and Odlyzko [3] that almost all the subset sum problem with density less than $0.6463\ldots$ would be solved in polynomial time with a single call to an oracle that can find the shortest vector in a special lattice. The bound was improved later to $0.9408\ldots$ by Coster *et al.* [1]. Li and Ma [4] gave a similar result when $d < 0.488\ldots$ for the extended versions of the subset sum problem where the variables are in $\{-1, 0, 1\}$ instead of $\{0, 1\}$ and the weights are allowed to be negative. Wang *et al.* [8] showed the same bound $d < 0.488\ldots$ holds for the extended modular subset problems.

In this paper, we discuss the density of the multiple subset sum problem and its some extended versions when given a lattice oracle. Given k subset sum problems with the **same** solution $(x_1, x_2, \cdots, x_n) \in \{0, 1\}^n$

$$\begin{array}{rcl} \sum\limits_{i=1}^{n} a_{1,i}x_i & = & s_1 \\ \sum\limits_{i=1}^{n} a_{2,i}x_i & = & s_2 \\ \vdots & \vdots & \vdots \\ \sum\limits_{i=1}^{n} a_{k,i}x_i & = & s_k, \end{array}$$

the multiple subset sum problem is to find the solution. Obviously, when $k = 1$, it agrees with the general subset sum problem. Similarly, we can define the density of the multiple subset sum problem as

$$d = \frac{n}{k \cdot \log(\max_{j,i} a_{ji})}.$$

As we know, Liu *et al.* [5] transformed the multiple subset sum problem to a new single subset sum problem, whose density is approximately equal to the density of the multiple subset sum problem we have defined. For the new single subset sum problem, they got the bound $d < 0.9408\ldots$ by the known result. However, just a heuristic explanation, but not a rigorous proof, was given for their result in [5]. Furthermore, it seems hard to apply their method for the extended versions of the modular subset sum problem.

In this paper, we give a rigorous proof for the result and generalize it to some extended versions of the problem, including the multiple modular subset sum problem. What's more, a modified lattice involved in our proof makes the analysis much simpler than before.

2 Preliminaries

We denote by $\mathbb{Z}$ the integer ring. We use bold letters to denote vectors, in row notation. If $\mathbf{v}$ is a vector, then we denote by v_i the i-th entry of $\mathbf{v}$. Let $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ be the Euclidean norm and inner product of $\mathbb{R}^n$.

2.1 Lattice

Let $B = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\} \subset \mathbb{R}^m$ be a set of n linearly independent vectors. The lattice $\mathcal{L}$ generated by the basis B is defined as

$$\mathcal{L}(B) = \{\sum_{i=1}^{n} x_i \mathbf{b}_i : x_i \in \mathbb{Z}\}.$$

n is called the dimension of the lattice. Finding a non-zero shortest vector of a lattice $\mathcal{L}$ is called the *shortest vector problem*(SVP).

2.2 The Number of Integer Points in $B_n(R)$

We denote by $B_n(R)$ the ball centered at the origin with radius R and by $N(n, R^2)$ the number of integer points in $B_n(R)$, i.e.,

$$N(n, R^2) = |\{\mathbf{z} \in \mathbb{Z}^n : \sum_{i=1}^{n} z_i^2 \le R^2\}|.$$

By the techniques of Mazo and Odlyzko [9], it can be shown that:

- $N(n, \frac{n}{2}) \le 2^{c_0 n}, c_0 = 1.54725...$ ([3]),
- $N(n, \frac{n}{4}) \le 2^{c_1 n}, c_1 = 1.0628...$ ([1]),
- $N(n, n) \le 2^{c_2 n}, c_2 = 2.047...$ ([4]).

2.3 The Density of the Subset Sum Problem

We present some results on the subset sum problem below.

Let

A be a positive integer,
$a_1, a_2, \ldots, a_n$ be random integers with $0 < a_i \le A$ for $1 \le i \le n$,
$\mathbf{e} = (e_1, e_2, \cdots, e_n)$ be any non-zero vector in $\{0,1\}^n$,
$s = \sum_{i=1}^{n} e_i a_i$.

Coster *et al.* [1] showed that if the density $d < 0.9408\ldots$, then the subset sum problem defined by $a_1, a_2, \ldots, a_n$ and s may "almost always" be solved in polynomial time with a single call to a lattice oracle which can solve the SVP. We sketch their proof here. Consider the lattice $\mathcal{L}$ generated by the basis

$$\begin{array}{rcl} \mathbf{b}_1 & = & (1, 0, \cdots, 0, Na_1) \\ \mathbf{b}_2 & = & (0, 1, \cdots, 0, Na_2) \\ & \vdots & \\ \mathbf{b}_n & = & (0, 0, \cdots, 1, Na_n) \\ \mathbf{b}_{n+1} & = & (\frac{1}{2}, \frac{1}{2}, \cdots, \frac{1}{2}, Ns), \end{array}$$

where N is an integer greater than $\sqrt{\frac{n}{4}}$. Notice that there is a vector $\bar{\mathbf{e}} = (e_1 - \frac{1}{2}, e_2 - \frac{1}{2}, \cdots, e_n - \frac{1}{2}, 0)$ in $\mathcal{L}$. If $\mathcal{L}$ contains no other non-zero vectors shorter than $\bar{\mathbf{e}}$ or $-\bar{\mathbf{e}}$, then we immediately obtain the vector $\bar{\mathbf{e}}$, hence the solution $\mathbf{e}$, with a single call to a lattice oracle. What's more, Coster *et al.* [1] showed the probability that $\mathcal{L}$ contains no other non-zero vectors shorter than $\bar{\mathbf{e}}$ or $-\bar{\mathbf{e}}$ is $1 - P$, where

$$P \leq n(4n\sqrt{n} + 1)\frac{2^{c_1 n}}{A}.$$

Hence, when n is large enough, if the density d is less than $\frac{1}{c_1}$ ($\approx 0.9408...$), P can be as small as possible, and with probability very near one, $\bar{\mathbf{e}}$ and $-\bar{\mathbf{e}}$ are the only shortest vectors in $\mathcal{L}$.

Li and Ma [4] extended the variables range from $\{0, 1\}$ to $\{-1, 0, 1\}$, allowed the weight to be negative. They also estimated the upper bound of the corresponding probability P and got

$$P \leq n(n + \frac{1}{2A^t})\frac{2^{c_2 n}}{A^{1-t}} + \frac{2}{A^t}$$

where $0 < t < 1$. Similarly, they showed for large enough n and $d < 0.488...$, P would be small enough.

The extended modular subset sum problem refers to the question to find variables $(x_1, x_2, \cdots, x_n) \in \{-1, 0, 1\}^n$, given positive integers $a_1, a_2, \ldots, a_n$, q and s, such that

$$\sum_{i=1}^{n} a_i x_i \equiv s (\text{mod } q).$$

Wang *et al.* showed the corresponding P satisfied

$$P \leq \frac{2^{c_2(n+1)}}{q} n(2n + 3)^2,$$

and $d < 0.488$.

2.4 The Multiple Subset Sum Problem

In this paper, we consider the multiple subset sum problem(Multiple SSP) and the corresponding multiple modular subset sum problem (Multiple Modular SSP).

Given a positive integer A, the multiple subset sum problem refers the question to recover $(x_1, x_2, \cdots, x_n) \in \{0,1\}^n$ from $a_{ji}(1 \leq j \leq k, 1 \leq i \leq n)$ and $s_1, s_2, \ldots, s_k$ where a_{ij}'s are uniformly independently and randomly chosen from the set of integers between 1 and A and $s_1, s_2, \ldots, s_k$ satisfy

$$\begin{array}{rcl} \sum_{i=1}^{n} a_{1,i}x_i & = & s_1 \\ \sum_{i=1}^{n} a_{2,i}x_i & = & s_2 \\ \vdots & \vdots & \vdots \\ \sum_{i=1}^{n} a_{k,i}x_i & = & s_k. \end{array}$$

The density of the multiple subset sum is defined by

$$d = \frac{n}{k \log_2(\max_{j,i} a_{ji})}.$$

Similarly, given a positive integer q, the multiple modular subset sum problem is to find $(x_1, x_2, \cdots, x_n) \in \{0,1\}^n$ from $a_{ji}(1 \leq j \leq k, 1 \leq i \leq n)$ and $s_1, s_2, \ldots, s_k$, where a_{ij}'s are uniformly independently and randomly chosen from the set of integers between 1 and $q-1$ and $s_1, s_2, \ldots, s_k$ satisfy

$$\begin{array}{rcl} \sum_{i=1}^{n} a_{1,i}x_i & \equiv & s_1 \pmod q \\ \sum_{i=1}^{n} a_{2,i}x_i & \equiv & s_2 \pmod q \\ \vdots & \vdots & \vdots \\ \sum_{i=1}^{n} a_{k,i}x_i & \equiv & s_k \pmod q. \end{array}$$

The density of the multiple modular subset sum is defined by

$$d = \frac{n}{k \log_2 q}.$$

3 A Note on the Density of the Multiple Subset Sum Problem

3.1 Multiple Subset Sum Problem

We give the main result for the general multiple subset sum problem first.

Theorem 1. *Let A be a positive integer, and let $a_{ji}(1 \leq j \leq k, 1 \leq i \leq n)$ be independently uniformly random integers between 1 and A. Let $\boldsymbol{e} = (e_1, e_2, \cdots, e_n)$ be arbitrary non-zero vector in $\{0,1\}^n$, and let $s_1 = \sum_{i=1}^n a_{1i}e_i, s_2 = \sum_{i=1}^n a_{2i}e_i, \ldots, s_k = \sum_{i=1}^n a_{ki}e_i$. If the density $d < 0.9408\ldots$, then the multiple subset sum problem defined by $a_{ji}(1 \leq j \leq k, 1 \leq i \leq n)$ and $s_1, s_2, \ldots, s_k$ may "almost always" be solved in polynomial time with a single call to a lattice oracle.*

Proof. Define vectors $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n, \mathbf{b}_{n+1}$ as follows:

$$\begin{array}{rcl}
\mathbf{b}_1 & = & (1, 0, \cdots, 0, 0, Na_{1,1}, Na_{2,1}, \cdots, Na_{k,1}) \\
\mathbf{b}_2 & = & (0, 1, \cdots, 0, 0, Na_{1,2}, Na_{2,2}, \cdots, Na_{k,2}) \\
& \vdots & \\
\mathbf{b}_n & = & (0, 0, \cdots, 1, 0, Na_{1,n}, Na_{2,n}, \cdots, Na_{k,n}) \\
\mathbf{b}_{n+1} & = & (\frac{1}{2}, \frac{1}{2}, \cdots, \frac{1}{2}, \frac{1}{2}, Ns_1, \ \ Ns_2, \ \ \cdots, \ \ Ns_k)
\end{array}$$

where N is an integer greater than $\sqrt{\frac{n+1}{4}}$.

Let $\mathcal{L}$ be the lattice generated by $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n, \mathbf{b}_{n+1}$. Then we can easily know that $\bar{\mathbf{e}} = (e_1 - \frac{1}{2}, e_2 - \frac{1}{2}, \cdots, e_n - \frac{1}{2}, -\frac{1}{2}, 0, \cdots, 0)$ is in $\mathcal{L}$. Notice that $\|\bar{\mathbf{e}}\| = \sqrt{\frac{n+1}{4}}$.

Let $X = \{\mathbf{v} \in \mathcal{L} | 0 < \|\mathbf{v}\| \leq \|\bar{\mathbf{e}}\|, \mathbf{v} \notin \{\mathbf{0}, \bar{\mathbf{e}}, -\bar{\mathbf{e}}\}\}$. If $X = \emptyset$, Then $\bar{\mathbf{e}}, -\bar{\mathbf{e}}$ are the only two non-zero shortest lattice vectors of $\mathcal{L}$. So we are interested in the probability that $X = \emptyset$. We first estimate the value of $Pr[X \neq \emptyset]$. Setting $N > \sqrt{\frac{n+1}{4}}$ implies that $v_{n+2} = v_{n+3} = \ldots = v_{n+1+k} = 0$ for any $\mathbf{v} \in X$. Suppose that $\mathbf{v} = \sum_{i=1}^{n+1} x_i \mathbf{b}_i \in X$, then we can express v_i in term of x_i in the following way

$$\begin{array}{rcll}
v_i & = & x_i + \frac{1}{2}x_{n+1} & i = 1, \ldots, n, \\
v_{n+1} & = & \frac{1}{2}x_{n+1}, & \\
v_{n+1+j} & = & N \cdot \{\sum_{i=1}^n a_{ji}x_i + x_{n+1}s_j\} = 0 & j = 1, \ldots, k.
\end{array}$$

This implies that

$$\sum_{i=1}^n a_{ji}(v_i - v_{n+1}) + 2v_{n+1}s_j = 0, j = 1, \ldots, k.$$

Let $\bar{\mathbf{v}} = (v_1, v_2, \ldots, v_{n+1})$. Then

$$
\begin{aligned}
Pr[X \neq \emptyset] \leq Pr[\exists \bar{\mathbf{v}}, s.t. 0 < \|\bar{\mathbf{v}}\| \leq \|\bar{\mathbf{e}}\|, \mathbf{v} \notin \{\mathbf{0}, \bar{\mathbf{e}}, -\bar{\mathbf{e}}\},& \\
\sum_{i=1}^{n} a_{ji}(v_i - v_{n+1}) + 2v_{n+1}s_j = 0, j = 1, \ldots, k]& \\
\leq Pr[\sum_{i=1}^{n} a_{ji}(v_i - v_{n+1}) + 2v_{n+1}s_j = 0, j = 1, \ldots, k, \mathbf{v} \notin \{\mathbf{0}, \bar{\mathbf{e}}, -\bar{\mathbf{e}}\}]\cdot& \\
|\{\bar{\mathbf{v}} | \|\bar{\mathbf{v}}\| \leq \|\bar{\mathbf{e}}\| = \sqrt{\frac{n+1}{4}}\}|.&
\end{aligned}
$$

For the second factor of the above expression, using the technique in [1] we know that, for large n,

$$
|\{\bar{\mathbf{v}} | \|\bar{\mathbf{v}}\| \leq \|\bar{\mathbf{e}}\| = \sqrt{\frac{n+1}{4}}\}| = N(n+1, \frac{n+1}{4}) + 2^{n+1} \leq 2^{c_1(n+1)}.
$$

Now we consider the first factor of *the above equation*. For $j = 1, \ldots, k$, since $s_j = \sum_{i=1}^{n} a_{ji}e_i$, we rewrite $\sum_{i=1}^{n} a_{ji}(v_i - v_{n+1}) + 2v_{n+1}s_j = 0$ as

$$
\sum_{i=1}^{n} a_{ji}z_i = 0, \text{ where } z_i = v_i - v_{n+1} + 2v_{n+1}e_i = x_i + x_{n+1}e_i.
$$

We claim that there must exist $t, 1 \leq t \leq n$, such that $z_t \neq 0$. Otherwise, $v_i = v_{n+1}(1-2e_i) = \pm v_{n+1}$, since $e_i \in \{0,1\}$, for $i = 1, \ldots, n$. Thus $\|\bar{\mathbf{v}}\| = \sqrt{n+1}|v_{n+1}| = \frac{\sqrt{n+1}}{2}|x_{n+1}|$ since $v_{n+1} = \frac{1}{2}x_{n+1}$. By the fact $\bar{\mathbf{v}} \in X$, we know that

$$
\|\bar{\mathbf{v}}\| = \frac{\sqrt{n+1}}{2}|x_{n+1}| \leq \frac{\sqrt{n+1}}{2},
$$

which implies that $|x_{n+1}| \leq 1$. Since $x_{n+1} \in \mathbb{Z}, x_{n+1}$ takes value only$-1, 0, 1$, corresponding to $\bar{\mathbf{v}} = \bar{\mathbf{e}}, \mathbf{0}, -\bar{\mathbf{e}}$, which contradicts to the definition of X. Thus,

$$\begin{aligned}
& Pr[\sum_{i=1}^{n} a_{ji}(v_i - v_{n+1}) + 2v_{n+1}s_j = 0, j = 1, \ldots, k, \mathbf{v} \notin \{\mathbf{0}, \bar{\mathbf{e}}, -\bar{\mathbf{e}}\}] \\
\leq\ & Pr[\sum_{i=1}^{n} a_{ji}z_i = 0, j = 1, \ldots, k] \\
=\ & Pr[a_{jt} = -\frac{\sum_{i=1, i\neq t}^{n} a_{ji}z_i}{z_t}, j = 1, \ldots, k] \\
=\ & \prod_{j=1}^{k} Pr[a_{jt} = -\frac{\sum_{i=1, i\neq t}^{n} a_{ji}z_i}{z_t}] \\
\leq\ & \frac{1}{A^k}.
\end{aligned}$$

Thus,

$$Pr[X \neq \emptyset] \leq \frac{2^{c_1 n}}{A^k} 2^{c_1}.$$

For large enough n, if $d < \frac{1}{c_1}$, it can be easily concluded that the probability of the event X is not empty can be very small, exponentially close to zero.

Thus, for large enough n, almost all multiple SSP with density $d < \frac{1}{c_1} = 0.9408\ldots$ can be solved in polynomial time with a single call to a lattice oracle.

□

Remark 1. *Notice that for $k = 1$, it is the general subset sum problem. The analysis is much simpler than before and the upper bound for $Pr[X \neq \emptyset]$ is also better and simpler than the previous results.*

If we also extend the extended the variables range from $\{0, 1\}$ to $\{-1, 0, 1\}$, allowed the weight to be negative, a similar result holds when $d < \frac{1}{c_2} = 0.488\ldots$ by considering the lattice generated by

$$\begin{array}{rcl}
\mathbf{b}_1 & = & (1, 0, \cdots, 0, 0, Na_{1,1}, Na_{2,1}, \cdots, Na_{k,1}) \\
\mathbf{b}_2 & = & (0, 1, \cdots, 0, 0, Na_{1,2}, Na_{2,2}, \cdots, Na_{k,2}) \\
 & \vdots & \\
\mathbf{b}_n & = & (0, 0, \cdots, 1, 0, Na_{1,n}, Na_{2,n}, \cdots, Na_{k,n}) \\
\mathbf{b}_{n+1} & = & (0, 0, \cdots, 0, 1, \ Ns_1, \ Ns_2, \ \cdots, \ Ns_k)
\end{array}$$

where N is an integer greater than $\sqrt{\frac{n+1}{4}}$.

3.2 Multiple Modular Subset Sum Problems

Theorem 2. *Let q be a positive integer greater than $\sqrt{\frac{n+1}{4}}$, and let $a_{ji}(1 \leq j \leq k, 1 \leq i \leq n)$ be independently uniformly random integers between 1 and $q-1$. Let $\boldsymbol{e} = (e_1, e_2, \cdots e_n)$ be arbitrary non-zero vector in $\{0,1\}^n$, and let $\sum_{i=1}^{n} a_{1i}e_i \equiv s_1 (\mathrm{mod}\ q)$, $\sum_{i=1}^{n} a_{2i}e_i \equiv s_2 (\mathrm{mod}\ q)$, $\ldots$ $\sum_{i=1}^{n} a_{ki}e_i \equiv s_k (\mathrm{mod}\ q)$, then with probability greater than 1- $\frac{2^{c_1 n}}{q^k} 2^{c_1}((n+1)\sqrt{n}+1)^k$, the multiple modular subset sum problem defined by $a_{ji}(1 \leq j \leq k, 1 \leq i \leq n)$ and $s_1, s_2, \ldots, s_k$ can be solved in polynomial time with a single call to a lattice oracle.*

Before giving the proof of Theorem 2, we first give two obvious corollaries.

Corollary 1. *For fixed k, If n is large enough and the density $d < 0.9408\ldots$, almost all the multiple modular subset sum problem defined by $a_{ji}(1 \leq j \leq k, 1 \leq i \leq n)$ and $s_1, s_2, \ldots, s_k$ can be solved in polynomial time with a single call to a lattice oracle.*

Corollary 2. *If $q > (n+1)\sqrt{n}+1$, the probability can be increased by increasing k, furthermore, if n is large enough and the density $d < 0.9408\ldots$, almost all the multiple modular subset sum problem defined by $a_{ji}(1 \leq j \leq k, 1 \leq i \leq n)$ and $s_1, s_2, \ldots, s_k$ can be solved in polynomial time with a single call to a lattice oracle.*

Proof. The method is similar to the proof of Theorem 1. Define the vectors as follows:

$$\begin{array}{lcl}
\mathbf{b}'_1 & = & (1,0,\cdots,0,0,Na_{1,1},Na_{2,1},\cdots,Na_{k,1}) \\
\mathbf{b}'_2 & = & (0,1,\cdots,0,0,Na_{1,2},Na_{2,2},\cdots,Na_{k,2}) \\
 & \vdots & \\
\mathbf{b}'_n & = & (0,0,\cdots,1,0,Na_{1,n},Na_{2,n},\cdots,Na_{k,n}) \\
\mathbf{b}'_{n+1} & = & (0,0,\cdots,0,0,\ \ Nq,\ \ \ 0,\ \ \ \cdots,\ \ \ 0\ \ ) \\
\mathbf{b}'_{n+2} & = & (0,0,\cdots,0,0,\ \ 0,\ \ \ Nq,\ \ \ \cdots,\ \ \ 0\ \ ) \\
 & \vdots & \\
\mathbf{b}'_{n+k} & = & (0,0,\cdots,0,0,\ \ 0,\ \ \ 0,\ \ \ \cdots,\ \ \ Nq\ ) \\
\mathbf{b}'_{n+k+1} & = & (\frac{1}{2},\frac{1}{2},\cdots,\frac{1}{2},\frac{1}{2},Ns_1,\ \ Ns_2,\ \ \cdots,\ \ Ns_k),
\end{array}$$

where $N > \sqrt{\frac{n+1}{4}}$, and consider the lattice $\mathcal{L}'$generated by the basis $\mathbf{b}'_1, \mathbf{b}'_2, \ldots, \mathbf{b}'_{n+k+1}$.

$$\begin{aligned}
Pr[X \neq \emptyset] \leq\ & Pr[\exists \bar{\mathbf{v}}, s.t. 0 < \|\bar{\mathbf{v}}\| \leq \|\bar{\mathbf{e}}\|, \mathbf{v} \notin \{\mathbf{0}, \bar{\mathbf{e}}, -\bar{\mathbf{e}}\}, \\
& \sum_{i=1}^{n} a_{ji}(v_i - v_{n+1}) + 2v_{n+1}s_j \equiv 0(\mathrm{mod}\ q), j = 1, \ldots, k] \\
=\ & Pr[\exists \bar{\mathbf{v}}, y_1, y_2, \ldots, y_k \in \mathbb{Z}, s.t. 0 < \|\bar{\mathbf{v}}\| \leq \|\bar{\mathbf{e}}\|, \mathbf{v} \notin \{\mathbf{0}, \bar{\mathbf{e}}, -\bar{\mathbf{e}}\}, \\
& \sum_{i=1}^{n} a_{ji}(x_i + e_i x_{n+1}) = q y_j, j = 1, \ldots, k] \\
\leq\ & |\{\bar{\mathbf{v}} | \|\bar{\mathbf{v}}\| \leq \|\bar{\mathbf{e}}\| = \sqrt{\frac{n+1}{4}}\}| \cdot \prod_{j=1}^{k} |\{y_j\}| \cdot \prod_{j=1}^{k} Pr[\sum_{i=1}^{n} a_{ji} z_i = q y_j]
\end{aligned}$$

For the first factor, it is clear that $|\{\bar{\mathbf{v}} | \|\bar{\mathbf{v}}\| \leq \|\bar{\mathbf{e}}\| = \sqrt{\frac{n+1}{4}}\}| \leq 2^{c_1(n+1)}$ for large n. Now we consider the second factor. Let $g_j = \frac{a_{ji}}{q}, j = 1, \ldots k$, then $|g_j| \leq 1, j = 1, \ldots k$.

$$\begin{aligned}
|y_j| &= |\sum_{i=1}^{n} g_i(v_i - v_{n+1} + 2v_{n+1}e_i)| \\
&= |\sum_{i=1}^{n} g_i v_i + \sum_{i=1}^{n}(2e_i - 1)v_{n+1}g_i| \\
&\leq \sum_{i=1}^{n} |g_i||v_i| + \sum_{i=1}^{n} |g_i||v_{n+1}| \\
&= (|v_1|, \ldots, |v_{n+1}|) \begin{pmatrix} |g_1| \\ \vdots \\ |g_n| \\ \sum\limits_{i=1}^{n} |g_i| \end{pmatrix} \\
&\leq \|v\| \cdot \sqrt{n + n^2} \\
&= \sqrt{\frac{n+1}{4}} \sqrt{n(n+1)} \\
&= \frac{n+1}{2} \sqrt{n}.
\end{aligned}$$

Thus,we have $\prod_{j=1}^{k} |\{y_j\}| \leq ((n+1)\sqrt{n} + 1)^k$.

Finally, for the last factor, we can previously conclude that for $\sum_{i=1}^{n} a_{ji} z_i = q y_j, 1 \leq j \leq k, \exists t, 1 \leq t \leq k,$such that $z_t \neq 0$, and, $\prod_{j=1}^{k} Pr[\sum_{i=1}^{n} a_{ji} z_i = q y_j] \leq \frac{1}{q^k}$.

Thus,

$$Pr[X \neq \emptyset] \leq \frac{2^{c_1 n}}{q^k} 2^{c_1}((n+1)\sqrt{n}+1)^k.$$

□

If q is a prime larger than $\sqrt{\frac{n+1}{4}}$, then we can get a better result.

Theorem 3. *Let q be a positive prime greater than $\sqrt{\frac{n+1}{4}}$, and let $a_{ji}(1 \leq j \leq k, 1 \leq i \leq n)$ be independently uniformly random integers between 1 and $q-1$. Let $\boldsymbol{e} = (e_1, e_2, \dots e_n)$ be arbitrary non-zero vector in $\{0,1\}^n$, and let $\sum_{i=1}^n a_{1i}e_i \equiv s_1 (\mathrm{mod}\ q)$, $\sum_{i=1}^n a_{2i}e_i \equiv s_2 (\mathrm{mod}\ q)$, $\dots$ $\sum_{i=1}^n a_{ki}e_i \equiv s_k (\mathrm{mod}\ q)$. If the density $d < 0.9408\dots$, then the multiple subset sum problem defined by $a_{ji}(1 \leq j \leq k, 1 \leq i \leq n)$ and $s_1, s_2, \dots, s_k$ may "almost always" be solved in polynomial time with a single call to a lattice oracle.*

Proof. The proof is similar to the one for Theorem 2 except that we can prove that there must exist $t, 1 \leq t \leq k$ such that $(z_t, q) = 1$ when q is a prime. Hence,

$$Pr[X \neq \emptyset] \leq \frac{2^{c_1 n}}{q^k} 2^{c_1}.$$

□

Remark 2. *Similarly, If $x_i \in \{-1, 0, 1\}$, then we can get*

$$Pr[X \neq \emptyset] \leq \frac{2^{c_2 n}}{q^k} 2^{c_2}((n+1)\sqrt{n}+1)^k,$$

when q is a prime larger than $\sqrt{\frac{1}{n+1}}$, the probability is

$$Pr[X \neq \emptyset] \leq \frac{2^{c_2 n}}{q^k} 2^{c_2}.$$

4 Algorithms to Solve the Lattice Problems

In this section,we give the algorithms to solve the shortest non-zero vector of the corresponding lattice problems of the multiple SSP and multiple modular SSP. Firstly, it is equivalent to solve the shortest non-zero vector of the intersection of several lattices. Furthermore, solving the multiple subset sum problem amouts to finding a small solution of an inhomogeneous linear equation, which can be viewed as a closest vector problem, by considering the corresponding homogeneous liear equation, together with an arbinary solution of the inhomogeneous equation.

4.1 Intersecting Lattices

For the multiple SSP, we should determine the non-zero shortest vector of the lattice $\mathcal{L}$. The row matrix is as follows:

$$\begin{bmatrix} 1,0,\ldots,0,0,Na_{1,1},Na_{2,1},\ldots,Na_{k,1} \\ 0,1,\ldots,0,0,Na_{1,2},Na_{2,2},\ldots,Na_{k,2} \\ \vdots \qquad\qquad\qquad\qquad \vdots \\ 0,0,\ldots,1,0,Na_{1,n},Na_{2,n},\ldots,Na_{k,n} \\ \frac{1}{2},\frac{1}{2},\ldots,\frac{1}{2},\frac{1}{2},Ns_1, \;\; Ns_2, \;\; \ldots, \;\; Ns_k \end{bmatrix}$$

Solving the shortest non-zero vector of the lattice $\mathcal{L}$ is equavilent to solving the shortest non-zero vector of the lattice

$$\mathcal{L}_1 \cap \mathcal{L}_2 \cap \cdots \cap \mathcal{L}_k$$

where $\mathcal{L}_i$ is the lattice generated by the row matirx

$$\begin{bmatrix} 1,0,\ldots,0,0,Na_{i,1} \\ 0,1,\ldots,0,0,Na_{i,2} \\ \vdots \quad \ddots \quad \vdots \quad \vdots \\ 0,0,\ldots,1,0,Na_{i,n} \\ \frac{1}{2},\frac{1}{2},\ldots,\frac{1}{2},\frac{1}{2},Ns_i \end{bmatrix}$$

$i = 1,\ldots,k$.

4.2 Kernel-Lattice

Assume the multiple subset sum problem

$$\begin{array}{rcl} \sum_{i=1}^{n} a_{1,i}x_i & = & s_1 \\ \sum_{i=1}^{n} a_{2,i}x_i & = & s_2 \\ \vdots & \vdots & \vdots \\ \sum_{i=1}^{n} a_{k,i}x_i & = & s_k \end{array}$$

has solution $\mathbf{e} = (e_1, e_2, \ldots, e_n)$. Let L is the set of all integer solutions to the homogeneous equations

$$\begin{aligned}\sum_{i=1}^{n} a_{1,i}x_i &= 0\\ \sum_{i=1}^{n} a_{2,i}x_i &= 0\\ \vdots \quad &\vdots\ \vdots\\ \sum_{i=1}^{n} a_{k,i}x_i &= 0\end{aligned}$$

L is a subgroup of $\mathbb{Z}^n$ and hence is a lattice. Let $y_1, \ldots, y_n$ be an arbitrary solution of the inhomogenous equation

$$\begin{aligned}\sum_{i=1}^{n} a_{1,i}x_i &= s_1\\ \sum_{i=1}^{n} a_{2,i}x_i &= s_2\\ \vdots \quad &\vdots\ \vdots\\ \sum_{i=1}^{n} a_{k,i}x_i &= s_k\end{aligned}$$

Then the vector $\mathbf{v} = (y_1 - e_1, \ldots, y_n - e_n) \in L$. And this vector is very close to the vector $\mathbf{t} = (y_1, \ldots, y_n)$. Thus by finding the closet vector to $\mathbf{t}$ in the lattice L, we may recover $\mathbf{v}$ and hence $\mathbf{e}$. The idea was also discussed in [7].

5 Conclusion

In this paper, we give some relations between the density of the multiple subset sum problem and the shortest vector in its corresponding lattice. Some extended versions are also considered. In addition, a modified lattice is involved to make the analysis much simpler than before.

References

[1] Coster, M.J. , Joux, A., LaMacchia, B.A., Odlyzko, A.M., Schnorr, C.-P. and Stern, J.: Improved low-density subset sum algorithms. Comput. Complexity, 2, 111C-128, (1992)

[2] Garey, M.R., Johnson, D.s.: Computer and Intractability: A Guide to the Theory of NP-Completeness. W.H. Freeman and CO., San Francisco, (1979)

[3] Lagarias, J. C., Odlyzko, A. M.: Solving Low-Lensity Subset Sum Problems. J. Assoc. Comp. Mach. 32(1) 229–246, (1985)

[4] Li, D., Ma, S.: Two Notes on Low-Density Subset Sum Algorithm. In ISAAC'94, Du, D., Zhang, X. ed. LNCS, vol. 834, pp. 164–171, Springer-Verlag, (1994)

[5] Liu, M., Wang, X., Bi, J. and Zheng, X.: Finding Shortest Lattice Vector for Lattci with Gaps. http://eprint.iacr.org/2011/139.

[6] Merkle, R., Hellman, M.: Hiding Information and Signatures in Trapdoor Knapsacks. IEEE Transactions on Information Theory 24(5), 525–530 (1978)

[7] Nguyen, Phong Q., Stern, J.: Adapting Density Attacks to Low-Weight Knapsacks, In Asiacrypt 2005, Lee, P., ed. LNCS, vol.3788, pp. 41–58, Springer, (2005)

[8] Wang, H., Xiao, H., Xiao, G.: EMSSP and its lattice reduction analysis. J. of Xidian University, 27(5), 616–618, (2000)

[9] J.E.Mazo, A.M.Odlyzko, Lattice points in High-Dimensional Spheres. Monatsh. Math,.110:47-61,1990.