

On the Hardness of Subset Sum Problem

Jun Kogure^{1,2}, Noboru Kunihiro², and
Hirosuke Yamamoto²

¹ Fujitsu Laboratories Ltd.,
4-1-1 Kamikodanaka, Nakahara-ku, Kawasaki, 211-8588, Japan

² The University of Tokyo,
5-1-5 Kashiwanoha, Kashiwa-shi, Chiba, 277-8561, Japan

Abstract. The subset sum problem, which is often called as the knapsack problem, is known as an NP-hard problem, and there are several cryptosystems based on the problem. The low-density attack algorithm by Lagarias and Odlyzko and its variants solve the subset sum problem efficiently, when the “density” of the given problem is smaller than some threshold. In the definition of the density, knapsacks are usually assumed to be chosen uniformly at random from the same interval. However, in general subset sum problem case, this assumption may not always hold. Further, changing values of knapsacks with maximum value fixed can affect the success probability of problem solving algorithms. In this paper, we assume that knapsacks are chosen from different intervals, and make analysis of the effect on the success probability of above algorithms both theoretically and experimentally.

Keywords. subset sum problem, knapsack-based cryptosystem, low-density attack, lattice reduction

1 Introduction

When a set of positive integers $S = \{a_1, \dots, a_n\}$ ($a_i \neq a_j$) and a positive integer s are given, determining whether there exists a subset of S with its sum being s , i.e. finding a vector $e = (e_1, \dots, e_n) \in \{0, 1\}^n$ satisfying $\sum_{i=1}^n a_i e_i = s$, is called the *subset sum problem* (or the knapsack problem), and is known as an NP-hard problem in general (see, e.g., [4]). Lagarias-Odlyzko (LO algorithm) [7] and Brickell [1] independently found an algorithm that solves subset sum problems, using lattice reduction algorithm. Both methods almost always solve the problem in polynomial time if we assume a shortest vector oracle of a lattice and if the density of the subset sum problem is < 0.6463 , where the density d is defined by

$$d = n / (\log_2 \max_i a_i). \quad (1)$$

Or if we assume all a_i 's are chosen uniformly at random from the interval $(0, A]$ for some integer A , the density may be written as

$$d = n / (\log_2 A).$$

Coster, Joux, LaMacchia, Odlyzko, Schnorr, and Stern raised the critical density up to 0.9408 (CJLOSS algorithm) [2]. Since these algorithms are effective

against relatively low-density subset sum problems, they are sometimes called the “low-density attack”. However, in general density case, the subset sum problem is still hard. In the low-density attacks, the subset sum problem is reduced to the *Shortest Vector Problem* (SVP) of a lattice constructed from the given problem, and one or two SVP oracle calls are admitted. Although no polynomial-time algorithms that solve Shortest Vector problem are known, the polynomial-time algorithm by Lenstra, Lenstra & Lovász (LLL algorithm)[6] solves it with some approximation factor and works relatively better in practice than in theory. One can also use the block Korkine-Zolotarev(BKZ) algorithm [10] (as in [11]), which provides better approximation factor but may not work in polynomial-time, if its block length parameter gets larger.

There have been proposed several public key cryptosystems whose security is based on the difficulty of the subset sum problem. For example, Chor-Rivest proposed a cryptosystem that can use subset sum problems with relatively high densities [3]. Though the system was attacked by an algebraic approach [12], the attack may not be valid in general cases. Okamoto-Tanaka-Uchiyama proposed another cryptosystem OTU, in an attempt to resist adversaries that can run quantum computers [9].

In these cryptosystems the Hamming weight of solutions is bounded by βn for a small constant $\beta \leq 1/2$. In general case, we can take $\beta = 1/2$. In the case β is relatively small, Coster et al. give improvements on their CJLOSS algorithm, which we refer as CJLOSS+ algorithm in this paper.

In the definition of the density, knapsack a_i 's are assumed to be chosen uniformly at random from the same interval $(0, A]$. However, in general subset sum problems, this assumption may not always hold. Further, in the definition (1), the density of the problem is determined only by the maximum value of the knapsacks. Even if we have same maximum value of knapsacks, experiments show different success probabilities of problem solving algorithms, when values of other knapsacks vary. In this paper, we assume that a_i 's are chosen from different interval $(0, A_i]$'s respectively. We introduce another density d_{HM} under this assumption, and see its validity as a criterion for the hardness of the subset sum problem theoretically. We also make experiments of solving subset sum problem changing the bit length of the knapsacks and make analysis of the effect on the success probability.

In section 2 we briefly look over the previous works regarding as the problem solving algorithms using lattice reduction, and consider the bit length of knapsacks which motivated our work. In section 3, we consider the case in which knapsacks are chosen from different intervals respectively. We present theoretical results in this case, and also look into non-asymptotic case and analyze the success probability of problem solving algorithm through experiments. Finally we conclude and state our future works.

2 Previous Works and Motivation

In this section, we review the low-density attack by Lagarias-Odlyzko (LO algorithm) and an improvement by Coster et al. (CJLOSS/CJLOSS+ algorithm).

Then we give our attentions to the bit length of knapsacks. We see the effect on the success probability of CJLOSS+ algorithm, when we change the bit length of knapsacks.

2.1 Low-density Attack Algorithms

First we review the LO algorithm:

INPUT: a_1, \dots, a_n and s
 OUTPUT: $(e_1, \dots, e_n) \in \{0, 1\}^n$ s.t. $\sum_{i=1}^n a_i e_i = s$
 PROCEDURE:
 $N \leftarrow \lfloor \sqrt{n} \rfloor$
 invoke a shortest vector oracle with the following basis:
 $b_1 = (1, 0, \dots, 0, Na_1),$
 $b_2 = (0, 1, \dots, 0, Na_2),$
 \vdots
 $b_n = (0, 0, \dots, 1, Na_n),$
 $b_{n+1} = (0, 0, \dots, 0, Ns);$
 let $(e'_1, \dots, e'_n, e'_{n+1})$ be the return value;
if $\sum_{i=1}^n \pm a_i e'_i = s$ and $\pm e'_i \in \{0, 1\}$ for all $1 \leq i \leq n$ and $e'_{n+1} = 0$
then output $\pm(e'_1, \dots, e'_n)$ and halt;
else
 output “not found”
end

Theorem 1 ([7]). *Let A be a positive integer, and let a_1, \dots, a_n be random integers with $0 < a_i \leq A$ for $1 \leq i \leq n$. Let $e = (e_1, \dots, e_n) \in \{0, 1\}^n$ be arbitrary, and let $s = \sum_{i=1}^n e_i a_i$. If the density $d < 0.6463$, then LO algorithm “almost always” solves the subset sum problem defined by a_1, \dots, a_n and s .*

As we would like to assume that the number of i 's such that $e_i = 1$ is $\leq \frac{n}{2}$, we actually execute the procedure also for $s' = (\sum_{i=1}^n a_i) - s$.

In CJLOSS algorithm, N is replaced by $\lfloor \frac{1}{2} \sqrt{n} \rfloor$, and vector b_{n+1} is replaced by

$$\left(\frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2}, Ns\right).$$

Checking if $\sum_{i=1}^n \pm a_i e'_i = s$ and $\pm e'_i \in \{0, 1\}$ is replaced by checking if $\sum_{i=1}^n a_i (\pm e'_i + \frac{1}{2}) = s$ and $e'_i \in \{\frac{1}{2}, -\frac{1}{2}\}$, and the output is replaced by $(\pm e'_1 + \frac{1}{2}, \dots, \pm e'_n + \frac{1}{2})$.

We also have the following theorem.

Theorem 2 ([2]). *Let A be a positive integer, and let a_1, \dots, a_n be random integers with $0 < a_i \leq A$ for $1 \leq i \leq n$. Let $e = (e_1, \dots, e_n) \in \{0, 1\}^n$ be arbitrary, and let $s = \sum_{i=1}^n e_i a_i$. If the density $d < 0.9408$, then CJLOSS algorithm “almost always” solves the subset sum problem defined by a_1, \dots, a_n and s .*

In some cryptosystems such as the Chor-Rivest cryptosystem, the Hamming weight k of solutions is bounded by $k = \beta n$ for a small constant $\beta \leq 1/2$. Coster et al. remarked further improvement (CJLOSS+ algorithm) in such cases. In CJLOSS+ algorithm, N is replaced by $\lfloor \sqrt{\beta(1-\beta)n} \rfloor$, and vector b_{n+1} is replaced by

$$(\beta, \beta, \dots, \beta, Ns).$$

Checking if $\sum_{i=1}^n \pm a_i e'_i = s$ and $\pm e'_i \in \{0, 1\}$ is replaced by checking if $\sum_{i=1}^n a_i (\pm e'_i + \beta) = s$ and $e'_i \in \{\pm(1-\beta), \pm\beta\}$, and the output is replaced by $(\pm e'_1 + \beta, \dots, \pm e'_n + \beta)$.

2.2 Bit Length of Knapsacks

In the previous subsection, it is assumed that knapsack a_i 's are chosen uniformly at random from a unique interval $(0, A]$. In the definition of the density (1), it is determined only by the maximum value of given knapsacks.

$$d = n / (\log_2 \max_i a_i).$$

Even if the maximum value of knapsacks is fixed, changing other knapsacks may effect the success probability of the solving algorithms. We see this through experiments. We take $n = 60$ and the Hamming weight k of the solution is 6. We take a_i 's of bit length 60 or 40, change their ratio, and run CJLOSS+ algorithm 100 times for each ratio. When we fix the ratio, we choose different sets of a_i 's 100 times without changing the ratio. As a lattice reduction algorithm, we use block Korkine-Zolotarev algorithm with block length 20. The table 1 shows the success probability of CJLOSS+ algorithm in percentage.

Table 1. Success Probability of CJLOSS+ algorithm

No. of 40-bit a_i 's	No. of 60-bit a_i 's	Success(%)
60	0	56
55	5	86
50	10	87
45	15	98
40	20	100
35	25	100
30	30	99
25	35	100
20	40	100
15	45	100
10	50	100
5	55	100
0	60	100

Though the density of a_i 's are all 1 except the top row of the table, success probability varies, which indicates the definition of the density may not be fully appropriate.

Further, in general subset sum problem, all knapsacks are not necessarily chosen from a unique interval. In the next section, we will consider the case knapsacks are chosen from different intervals.

3 Results in case Knapsacks are from Different Intervals

In previous works, it is assumed that knapsacks are chosen from a unique interval. In this section we assume that they are chosen from different intervals respectively, and we describe our theoretical and experimental results in that case.

3.1 Theoretical Results in Asymptotic Case

Theorem 3. *Let $e = (e_1, \dots, e_n) \neq (0, \dots, 0) \in \{0, 1\}^n$ be fixed. Let A_1, \dots, A_n be positive integers with $A_1 \leq \dots \leq A_n$, and a_1, \dots, a_n be integers chosen uniformly and independently at random with $0 < a_i \leq A_i$ for $1 \leq i \leq n$. Let $s = \sum_{i=1}^n e_i a_i$, and let L be a lattice spanned by the following basis:*

$$\begin{aligned} b_1 &= (1, 0, \dots, 0, Na_1), \\ b_2 &= (0, 1, \dots, 0, Na_2), \\ &\vdots \\ b_n &= (0, 0, \dots, 1, Na_n), \\ b_{n+1} &= (0, 0, \dots, 0, Ns), \end{aligned}$$

where N is a positive integer larger than \sqrt{n} . Let $\delta(u_0)$ be the minimum value of the following function of $u \in \mathbb{R}^+$:

$$\delta(u) = \frac{1}{2}u + \ln \theta(e^{-u}), \quad \theta(z) = 1 + 2 \sum_{j=1}^{\infty} z^{j^2},$$

and let c_0 denote $(\log_2 e)\delta(u_0)$.

Then the probability P that the shortest vector in L is not equal to $\hat{e} = (e_1, \dots, e_n, 0)$ is less than

$$(2n\sqrt{n/2} + 1)2^{c_0 n} \sum_{i=1}^n \frac{1}{A_i}.$$

Note that the critical density $d_0 = 0.6463\dots$ in LO algorithm case coincides with $\frac{1}{c_0}$ in above statement[8].

The proof, we will give in the following, is based on the proof in [2].

Proof. Let $t = \sum_{i=1}^n a_i$. We may assume that

$$\frac{1}{n}t \leq s \leq \frac{n-1}{n}t,$$

because otherwise any $a_i > t/n$ may be removed from consideration. The vector $\hat{e} = (e_1, \dots, e_n, 0)$ is contained in L . We should consider the probability that there exists a vector $\hat{x} = (x_1, \dots, x_n, x_{n+1})$ satisfying the following conditions:

$$\|\hat{x}\| \leq \|\hat{e}\|, \quad \hat{x} \in L, \quad \hat{x} \notin \{0, \pm\hat{e}\}, \quad (2)$$

where $\|x\|$ represents Euclidean norm of x . Then \hat{x} satisfies the condition (2) only when $x_{n+1} = 0$, because otherwise we have $\|\hat{x}\| \geq |x_{n+1}| \geq N > \sqrt{n} \geq \|\hat{e}\|$ which contradicts the condition (2). Hence we have some integer y that satisfies

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

Then

$$|y| \leq n\sqrt{n/2}$$

holds, because

$$|y|s = \left| \sum_{i=1}^n x_i a_i \right| \leq \|\hat{x}\| \left| \sum_{i=1}^n a_i \right| = \|\hat{x}\|t,$$

and without loss of generality we may assume that $\|e\| \leq n/2$. Let x denote $x = (x_1, \dots, x_n)$, and $z_i = x_i - ye_i$. Then we have

$$P \leq \#\{x \in \mathbb{Z}^n \mid \|x\| \leq \|e\|\} \cdot \#\{y \in \mathbb{Z} \mid |y| \leq n\sqrt{n/2}\} \cdot \Pr \left[\sum_{i=1}^n a_i z_i = 0 \right]. \quad (3)$$

When $z_n = z_{n-1} = \dots = z_{i+1} = 0$ and $z_i \neq 0$, let z' denote $z' = -\frac{1}{z_i} \sum_{j=1}^{i-1} a_j z_j$, then

$$\begin{aligned} & \Pr \left[\sum_{j=1}^n a_j z_j = 0 \mid z_n = z_{n-1} = \dots = z_{i+1} = 0, z_i \neq 0 \right] \\ &= \Pr[a_i = z' \mid z_n = z_{n-1} = \dots = z_{i+1} = 0, z_i \neq 0] \\ &= \sum_{l=1}^{A_i} \Pr[a_i = l] \Pr[z' = l \mid z_n = z_{n-1} = \dots = z_{i+1} = 0, z_i \neq 0] \\ &= \frac{1}{A_i} \sum_{l=1}^{A_i} \Pr[z' = l \mid z_n = z_{n-1} = \dots = z_{i+1} = 0, z_i \neq 0] \\ &\leq \frac{1}{A_i}. \end{aligned}$$

Hence we can estimate the last term of (3) by

$$\begin{aligned}
\Pr \left[\sum_{j=1}^n a_j z_j = 0 \right] &= \sum_{i=1}^n \Pr[z_n = z_{n-1} = \dots = z_{i+1} = 0, z_i \neq 0] \\
&\quad \times \Pr[a_i = z' | z_n = z_{n-1} = \dots = z_{i+1} = 0, z_i \neq 0] \\
&\leq \sum_{i=1}^n \Pr[z_n = z_{n-1} = \dots = z_{i+1} = 0, z_i \neq 0] \frac{1}{A_i} \quad (4) \\
&\leq \sum_{i=1}^n \frac{1}{A_i}.
\end{aligned}$$

□

Corollary 1. Let $HM(A_1, \dots, A_n)$ denote the harmonic mean of A_1, \dots, A_n , i.e.

$$HM(A_1, \dots, A_n) = \frac{1}{\frac{1}{A_1} + \dots + \frac{1}{A_n}}.$$

If for some $c > c_0$,

$$\lim_{n \rightarrow \infty} \log_2 HM(A_1, \dots, A_n) = cn,$$

then

$$P \rightarrow 0(n \rightarrow \infty).$$

Proof. From (3), we have

$$\lim_{n \rightarrow \infty} P \leq \lim_{n \rightarrow \infty} \frac{n(2n\sqrt{n/2} + 1)2^{c_0 n}}{HM(A_1, \dots, A_n)} = 0.$$

□

Above corollary indicates that in case we choose a_i 's from different periods A_1, \dots, A_n , we may use another indicator as its density:

$$d_{HM} = \frac{n}{\log_2 HM(A_1, \dots, A_n)}.$$

d_{HM} can be regarded as a natural extension of usual density d , because if all A_i 's are the same value A , d_{HM} coincides with d . As we currently assume all A_i 's are not same, we cannot bound the left term of the inequality (4) by $\frac{1}{\max_i A_i} = \frac{1}{A_n}$ in the proof of the theorem 3. Instead we use the harmonic mean of A_i 's, and we have $\frac{1}{A_n} \leq \frac{1}{HM(A_1, \dots, A_n)} \leq \frac{1}{\min_i A_i} = \frac{1}{A_1}$.

Further, we may combine this density with Kunihiro's density [5] $D = \frac{nH(\frac{k}{n})}{\log_2 A}$, where H is an Entropy function $H(x) = -x \log x - (1-x) \log(1-x)$, and k is the Hamming weight of the solution:

$$D_{HM} = \frac{nH(\frac{k}{n})}{\log_2 HM(A_1, \dots, A_n)}.$$

In the case of CJLOSS algorithm, we similarly have following results:

Theorem 4. Let $e = (e_1, \dots, e_n) \neq (0, \dots, 0) \in \{0, 1\}^n$ be fixed. Let A_1, \dots, A_n be positive integers with $A_1 \leq \dots \leq A_n$, and a_1, \dots, a_n be integers chosen uniformly and independently at random with $0 < a_i \leq A_i$ for $1 \leq i \leq n$. Let $s = \sum_{i=1}^n e_i a_i$, and let L be a lattice spanned by the following basis:

$$\begin{aligned} b_1 &= (1, 0, \dots, 0, Na_1), \\ b_2 &= (0, 1, \dots, 0, Na_2), \\ &\vdots \\ b_n &= (0, 0, \dots, 1, Na_n), \\ b'_{n+1} &= \left(\frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2}, Ns\right), \end{aligned}$$

where N is a positive integer larger than $\frac{1}{2}\sqrt{n}$. Let $\delta_{\frac{1}{2}}(u_1)$ be the minimum value of the following function of $u \in \mathbb{R}^+$:

$$\delta_{\frac{1}{2}}(u) = \frac{1}{4}u + \ln \theta(e^{-u}), \quad \theta(z) = 1 + 2 \sum_{j=1}^{\infty} z^{j^2},$$

and let c_1 denote $(\log_2 e) \delta_{\frac{1}{2}}(u_1)$.

Then the probability P that the shortest vector in L is not equal to $\hat{e}' = (e_1 - \frac{1}{2}, \dots, e_n - \frac{1}{2}, 0)$ is less than

$$(4n\sqrt{n} + 1)2^{c_1 n} \sum_{i=1}^n \frac{1}{A_i}.$$

Note that the critical density $d_1 = 0.9408\dots$ in CJLOSS algorithm case coincides with $\frac{1}{c_1}$ in above statement.

Corollary 2. Let $HM(A_1, \dots, A_n)$ denote the harmonic mean of A_1, \dots, A_n . If for some $c > c_1$,

$$\lim_{n \rightarrow \infty} \log_2 HM(A_1, \dots, A_n) = cn,$$

then

$$P \rightarrow 0 (n \rightarrow \infty).$$

In case of CJLOSS+ algorithm, we use the following function $\delta_\beta(u)$ of $u \in \mathbb{R}^+$:

$$\delta_\beta(u) = \beta(1 - \beta)u + \ln \theta(e^{-u}), \quad \theta(z) = 1 + 2 \sum_{j=1}^{\infty} z^{j^2}.$$

If we set $\beta = \frac{1}{2}$, this function coincides with $\delta_{\frac{1}{2}}(u)$ in theorem 4.

3.2 In Non-Asymptotic case

Let P_i denote the probability $\Pr[z_n = z_{n-1} = \dots = z_{i+1} = 0, z_i \neq 0]$ in the inequality (4) of the proof of the theorem 3. In the proof, we bounded each term $P_i \frac{1}{A_i}$ in the inequality (4) with $\frac{1}{A_i}$ in an asymptotic case. However, when we deal with concrete subset sum problems, we should rather analyze the coefficients P_i minutely. For simplicity we first consider the case $y = 0$, i.e. $z_i = x_i$ for any i , to see the behavior of these values. As x satisfies $x \in \mathbb{Z}^n$ and $\|x\| \leq \|e\|$, the total number of candidates for x is $N(n, \|e\|^2)$, where $N(n, r^2)$ denotes the number of integral points in the sphere centered at the origin with radius r . The number of points that satisfy $x_n = x_{n-1} = \dots = x_{i+1} = 0$ is $N(i, r^2)$, and the number of points that satisfy $x_n = x_{n-1} = \dots = x_{i+1} = x_i = 0$ is $N(i-1, r^2)$. Hence we have

$$\Pr[z_n = z_{n-1} = \dots = z_{i+1} = 0, z_i \neq 0] = \frac{N(i, \|e\|^2) - N(i-1, \|e\|^2)}{N(n, \|e\|^2)}.$$

Table 2 shows the concrete values of these probabilities in percentage, when we take $n = 60, r^2 = 6$. We can count up the number of these points as follows: As we set $r^2 = 6$, x_i cannot be larger than 2 for any i . In the case all non-zero elements of x are 1, if the number of non-zero elements is $k (1 \leq k \leq 6)$, we have ${}_i C_k \times 2^k$ such points, considering the positions of non-zero elements and their signs. In the case we have one 2 in non-zero elements, we have three patterns: $(1, 1, 2), (1, 2), (2)$. Each pattern has ${}_i C_1 \times {}_{i-1} C_2 \times 2^3, {}_i C_1 \times {}_{i-1} C_1 \times 2^2, {}_i C_1 \times 2$ points respectively.

Table 2. Probability of $z_n = z_{n-1} = \dots = z_{i+1} = 0, z_i \neq 0$ in case $y = 0$

i	$N(i, 6)$	$N(i-1, 6)$	$N(i, 6) - N(i-1, 6)$	$P_i (\%)$
$i=60$	3387774105	3052225761	335548344	9.90468
55	1972973993	1760010013	212963980	6.28625
50	1089132021	959978405	129153616	3.81234
45	563230189	489138937	74091252	2.18702
40	268500497	228849609	39650888	1.17041
35	115324945	95918421	19406524	0.57284
30	43135533	34703373	8432160	0.2489
25	13314261	10212465	3101796	0.0915585
20	3093129	2203697	889432	0.0262542
15	454137	285069	169068	0.00499053
10	29285	14581	14704	0.000434031
5	573	233	340	0.0000100361

Further we need to consider the case $y \neq 0$. When $y \neq 0$, $z_i = x_i - y$ for i such that $e_i = 1$. In this case we see the number of points as follows.

$$\begin{aligned} \#\{z|z_n = z_{n-1} = \dots = z_{i+1} = 0\} &= N(i, r^2) \\ \#\{z|z_n = z_{n-1} = \dots = z_{i+1} = 0, z_i = y\} &= N(i-1, r^2 - y^2) \\ \#\{z|z_n = z_{n-1} = \dots = z_{i+1} = 0, z_i = y, z_{i-1} = 0\} &= N(i-2, r^2 - y^2) \end{aligned}$$

As $N(i-1, r^2 - y^2) < N(i-1, r^2)$, $N(i, r^2) - N(i-1, r^2 - y^2) > N(i, r^2) - N(i-1, r^2)$, and P_i gets larger if $e_i = 1$. On the other hand, Both $N(i-1, r^2 - y^2)$ and $N(i-2, r^2 - y^2)$ are much smaller than $N(i-1, r^2)$ and $N(i-2, r^2)$ respectively, P_{i+1} 's from then on are expected to get much smaller. Hence the effect of A_i 's will get further smaller as i gets smaller, than the case $y = 0$.

This gives an explanation of the experiments in the next subsection, where the effects of A_i 's for smaller i 's on the success probability are relatively smaller than expected at first sight.

3.3 Experimental Results

In this subsection we make analysis of the experiments we saw in subsection 2.2. We implemented the CJLOSS+ algorithm and saw its success probability changing the ratio of the bit length of the knapsacks.

As in the previous subsection, we took $n = 60$ and $r^2 = 6$, which means the Hamming weight k of the solution is 6. We took a_i 's of bit length 60 or 40, changed their ratio, and ran CJLOSS+ algorithm 100 times for each ratio. When we fixed the ratio, we chose different sets of a_i 's 100 times without changing the ratio. As a lattice reduction algorithm, we used block Korkine-Zolotarev algorithm with block length 20. Table 1 in subsection 2.2 shows the success probability of CJLOSS+ algorithm in percentage.

Although the value of 40-bit a_i is far greater than 60-bit a_i , its effect becomes small if the number of 40-bit a_i 's gets less than 30.

As we see in previous subsection, we should consider the effect of P_i 's. As

$$\frac{1}{2^{40}} \approx 10^6 \times \frac{1}{2^{60}},$$

the effect of 40-bit a_i will become the same level as that of 60-bit a_i , if $P_i \leq 10^{-6} \times P_{60}$. More strictly, as we have only two kinds of bit-length, we can consider the sum of these probabilities. If we could find k that satisfies

$$\left(\sum_{i=0}^k P_i\right) \times \frac{1}{2^{40}} \approx \left(\sum_{i=k+1}^{60} P_i\right) \times \frac{1}{2^{60}},$$

more than k 40-bit a_i 's will affect the total success probability. From table 2 in the previous section, we see

$$\frac{N(60, 6)}{N(10, 6)} \approx 1.15 \times 10^5.$$

Hence k seems to be a bit smaller than 10, but this table 2 is the one in case $y=0$. When $y \neq 0$, P_i will get smaller for $i < j$, each time there appears j such that $e_j = 1$. This happens 6 times in this case, because the Hamming weight of the solution is 6. This makes the actual value of k larger, which seems to be around 30 from the table 1.

Another factor we have to consider is the approximate factor of the actual lattice reduction algorithms. However, as the running time grows exponentially if we use the exact algorithms, considering this effect is a difficult task in analyzing results of actual experiments.

4 Concluding Remarks

In this paper, we considered the hardness of general subset sum problems with an assumption that knapsacks are chosen from different intervals respectively. In asymptotic case, we introduced another density that works as an criterion for the success probabilities of problem solving algorithms that use lattice reduction. In non-asymptotic case, we actually calculate the weight of the effect of each interval. We also see their relations with the success probability of solving algorithms through concrete experiments. In the cryptographic aspect, mixing public keys whose bit lengths are smaller than the number of public keys may strengthen the system. However, when we fix parameters for cryptographic use(in non-asymptotic case), we should make a closer look at actual values of the effect of each knapsack, in order to estimate the success probability of the LO algorithm and its variants.

Our future work will be to get tighter bounds for the success probability of LO algorithm and its variants, which will be useful for estimating the security of actual knapsack-based cryptosystems more precisely.

References

1. E. F. Brickell, "Breaking iterated knapsacks", In *Advances in Cryptology: Proceedings of CRYPTO'84*, LNCS 196, pp.342–358, Springer-Verlag, 1985.
2. M. J. Coster, A. Joux, B. A. LaMacchia, A. M. Odlyzko, C. P. Schnorr, and J. Stern, "Improved low-density subset sum algorithms", *Computational Complexity*, 2, pp.111–128, 1992.
3. B. Chor, and R. L. Rivest, "A knapsack-type public key cryptosystem based on arithmetic in finite fields", *IEEE Trans. Information Theory*, 34(5), pp.901–909, 1988.
4. M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman and Co., San Francisco, 1979.
5. N. Kunihiko, "New Definition of Density on Knapsack Cryptosystems", In *Progress in Cryptology: Proceedings of Africacrypt 2008*, LNCS 5023, pp.156–173, Springer, 2008.
6. A. K. Lenstra, H. W. Lenstra Jr., and L. Lovász, "Factoring polynomials with rational coefficients", *Mathematische Annalen*, 261, pp.515–534, 1982.

7. J. C. Lagarias, and A. M. Odlyzko, "Solving low-density subset sum problems", J. ACM, 32(1), pp. 229–246, 1985.
8. J. E. Mazo, and A. M. Odlyzko, "Lattice points in high-dimensional spheres", Monatsch. Math., 110, pp.47–61, 1990.
9. T. Okamoto, K. Tanaka, and S. Uchiyama, "Quantum public-key cryptosystems", In *Advances in Cryptology: Proceedings of CRYPTO 2000*, LNCS 1880, pp.147–165, Springer, 2000.
10. C. P. Schnorr, and M. Euchner, "Lattice basis reduction: Improved practical algorithms and solving subset sum problems", Mathematical Programming, 66, pp.181–199, 1994.
11. C. P. Schnorr and H. H. Hörner, "Attacking the Chor-Rivest cryptosystem by improved lattice reduction", In *Advances in Cryptology: Proceedings of EURO-CRYPT'95*, LNCS 921, pp.1–12, Springer, 1995.
12. S. Vaudenay, "Cryptanalysis of the Chor–Rivest cryptosystem", J. Cryptology, 14(2), pp.87–100, 2001.