PAPER

On the Hardness of Subset Sum Problem from Different Intervals

Jun KOGURE[†a)], **Noboru KUNIHIRO**[††], ***Members***, ***and*** **Hirosuke YAMAMOTO**[††], ***Fellow***

SUMMARY 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. Assuming an oracle for shortest vector problem of lattice, 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. When we define the density in the context of knapsack-type cryptosystems, weights are usually assumed to be chosen uniformly at random from the same interval. In this paper, we focus on general subset sum problems, where this assumption may not hold. We assume that weights are chosen from different intervals, and make analysis of the effect on the success probability of above algorithms both theoretically and experimentally. Possible application of our result in the context of knapsack cryptosystems is the security analysis when we reduce the data size of public keys.

key words: *subset sum problem, knapsack problem, low-density attack, lattice reduction*

1. Introduction

When a set of positive integers (weights) $S = \{a_1, \ldots, a_n\}$ $(a_i \neq a_j)$ and a positive integer s are given, finding a vector $e = (e_1, \ldots, e_n) \in \{0, 1\}^n$ 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 [8] and Brickell [1] independently found an algorithm (LO algorithm, hereafter) 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 less than $0.6463\ldots$, where the density d is defined by

$$d = n/(\log_2 \max_i a_i). \tag{1}$$

Coster, Joux, LaMacchia, Odlyzko, Schnorr, and Stern raised the critical density up to $0.9408\ldots$ (CJLOSS algorithm, hereafter) [2]. They assumed that all a_i's are chosen uniformly at random from an interval $(0, A]$ for some integer A, and the density was defined as

$$d = n/(\log_2 A). \tag{2}$$

Since these algorithms are effective against subset sum problems with relatively low densities, they are sometimes called the "low-density attack" in the context of breaking knapsack-type cryptosystems. However, in general density cases, the subset sum problem is still hard. In the LO algorithm, 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) [7] 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 [11] (as in [12]), 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 hardness 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 [13], 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 [10].

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

Our Motivation and Contributions:

In the context of knapsack-type cryptosystems, public key a_i's are often generated by taking the value mod A for some integer A. Hence it would be reasonable to adapt the following assumption:

Assumption 1. *a_i's are chosen uniformly at random from the same interval* $(0, A]$.

In this case, the density can be defined as Eq. (2), and the effectiveness of LO algorithm is well analyzed.

On the other hand, in general subset sum problems, this assumption may not always hold and the effectiveness of LO algorithm is not well known. In this paper, we focus on general subset sum problems and analyze its hardness, mainly from theoretical interests. As LO algorithm can be applied to general subset sum problems and often works efficiently, analyzing its effectiveness is very important in order to an-

Manuscript received September 26, 2011.

[†]The author is with Fujitsu Laboratories Ltd., Kawasaki-shi, 211-8588 Japan.

[††]The authors are with The University of Tokyo, Kashiwa-shi, 277-8561 Japan.

a) E-mail: kogure@jp.fujitsu.com

DOI: 10.1587/transfun.E95.A.903

Copyright © 2012 The Institute of Electronics, Information and Communication Engineers

alyze the hardness of general subset sum problems. In general cases, we are given a_i's without knowing from which interval they are chosen. Given a_i's, we may adapt the following assumption:

Assumption 2. *a_i's are chosen uniformly at random from the same interval* $(0, \max_i a_i]$.

If we take this assumption, we can define the density as Eq. (1). However, if the bit lengths of a_i's vary, this assumption is not appropriate because the expected bit length is around $\log_2(\max_i a_i) - 1$. Actually, even if we have same maximum value of weights, i.e. same density, experiments show different success probabilities of LO algorithm when bit lengths of other weights vary. We will see this phenomenon in the following section.

Another possible assumption will be:

Assumption 3. *a_i is chosen uniformly at random from the interval* $(0, 2^{\lceil \log_2 a_i + 1 \rceil} - 1]$.

As the expected bit length of an integer that is chosen uniformly at random from the interval $(0, 2^m - 1]$ is around $m-1$, this assumption would be reasonable in some sense. In a nutshell, this assumption means that small weight is chosen from a small interval and large weight is chosen from a large interval. Hence, in this paper we analyze the effectiveness of LO algorithm when we adapt this assumption 3.

In general cases, efficient attacks might be possible by decomposing the problem, but we focus on solving the problem by LO algorithm, as it can be applied in any case. 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 lengths of the weights and make analysis of the effect on the success probability.

Possible application of our work in the context of knapsack-type cryptosystems is the security analysis of the system when we reduce the data size of public keys. If we would like to reduce the total public key size in knapsack-type cryptosystems, we need to have weights with shorter bit length. In order to assure the security of such systems, we need to analyze the hardness of general subset sum problems in our setting.

In Sect. 2 we briefly look over the previous works regarding as LO algorithm and its variants using lattice reduction, and consider changing the bit lengths of weights which motivated our work. In Sect. 3, we assume that weights are chosen from different intervals respectively, and present theoretical results in asymptotic case. We also look into non-asymptotic case and analyze the success probability of LO algorithm and its variants through experiments.

2. Previous Works and Concerns

In this section, we review LO algorithm by Lagarias and Odlyzko, and improvements by Coster et al. (CJLOSS/CJLOSS+ algorithm).

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

2.1 LO Algorithm and its Variants

First we review the LO algorithm:

INPUT: $a_1, ..., a_n$ and s
OUTPUT: $(e_1, ..., e_n) \in \{0,1\}^n$ 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:

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

let $(e'_1, ..., e'_n, e'_{n+1})$ be the return value;
if $\sum_{i=1}^n \pm a_i e'_i = s$ and $\pm e'_i \in \{0,1\}$ for all $1 \le i \le n$ and $e'_{n+1} = 0$
then output $\pm(e'_1, ..., e'_n)$ and halt;
else
output "not found"
end

Theorem 1 ([8])**.** *Let A be a positive integer, and let $a_1, \ldots, a_n$ be random integers with $0 < a_i \le A$ for $1 \le i \le n$. Let $e = (e_1, \ldots, e_n) \in \{0,1\}^n$ be arbitrary, and let $s = \sum_{i=1}^n e_i a_i$. If the density $d < d_0 = 0.6463\ldots$, then* LO *algorithm "almost always" solves the subset sum problem defined by $a_1, \ldots, a_n$ and s, assuming a shortest vector problem oracle.*

As we would like to assume that the number of i's such that $e_i = 1$ is less than or equal to $\frac{n}{2}$, 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

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

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 $(\pm e'_1 + \frac{1}{2}, \ldots, \pm e'_n + \frac{1}{2})$. We also have the following theorem.

Theorem 2 ([2])**.** *Let A be a positive integer, and let $a_1, \ldots, a_n$ be random integers with $0 < a_i \le A$ for $1 \le i \le n$. Let $e = (e_1, \ldots, e_n) \in \{0,1\}^n$ be arbitrary, and let $s = \sum_{i=1}^n e_i a_i$. If the density $d < d_1 = 0.9408\ldots$, then* CJLOSS *algorithm "almost always" solves the subset sum problem defined by $a_1, \ldots, a_n$ and s, assuming a shortest vector problem oracle.*

Table 1 Success probability of CJLOSS+ algorithm (in case the ratio of two kinds of bit lengths varies).

No. of 40-bit a_i's	No. of 60-bit a_i's	Success(%)
60	0	60.0
59	1	72.3
55	5	85.1
50	10	88.0
45	15	98.0
40	20	100.0
35	25	100.0
30	30	99.9
25	35	99.7
20	40	100.0
15	45	100.0
10	50	100.0
5	55	99.9
0	60	100.0

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

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

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 $(\pm e'_1 + \beta, \ldots, \pm e'_n + \beta)$.

2.2 Changing Bit Lengths of Weights

In the definition of the density (1), it is determined only by the maximum value of given weights if the number n of weights is fixed.

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

Even if the maximum value of weights is fixed, changing the bit lengths of other weights may effect the success probability of LO algorithms and its variants. We see this through experiments.

First, 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 1000 times for each ratio. When we fix the ratio, we choose different sets of a_i's 1000 times without changing the ratio. As a lattice reduction algorithm, we use block Korkine-Zolotarev algorithm with block length 20. Table 1 shows the success probability of CJLOSS+ algorithm in percentage. Though the density of a_i's are almost 1 except the top row of the table, success probability varies.

We see another pattern of weights where bit lengths are uniformly distributed. The numbers in the left column of Table 2 represents the number n of weights. The numbers in the first row represents the bit length m of a_i's. "71 − 80" means that bit lengths are uniformly distributed from 71 to 80, i.e. there are 7 or 8 a_i's for each bit length. Other numbers in the table represent the success probability of CJLOSS+ algorithm in percentage, when we run the algorithm 1000 times, generating n random weights of bit length m for each time. As a lattice reduction algorithm, we use block Korkine-Zolotarev algorithm with block length 20. The column of bit length $m = 71 - 80$ and the column of bit length $m = 80$ in Table 2 has almost the same density according to the definition (1), as the maximum bit length m of weights is 80. Even though they have almost the same density, the success probability gets smaller for $m = 71-80$.

Table 2 Success probability of CJLOSS+ algorithm (in case bit lengths of weights vary).

	$m = 71$	71-80	74	80
$n = 70$	95.0	98.5	98.8	99.9
80	36.2	51.1	44.6	58.8

These phenomena indicate that the definition (1) of the density may not be fully appropriate.

3. Analysis in Case Weights are Chosen from Different Intervals

In previous works, it is assumed that weights 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 analysis in that case.

3.1 Theoretical Results in Asymptotic Case

Theorem 3. *Let $e = (e_1, \ldots, e_n) \neq (0, \ldots, 0) \in \{0, 1\}^n$ be fixed. Let $A_1, \ldots, A_n$ be positive integers and $a_1, \ldots, a_n$ 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:*

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

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 $\hat{e} = (e_1, \ldots, e_n, 0)$ 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 $d_0 = 0.6463\ldots$ in LO algorithm case coincides with $\frac{1}{c_0}$ in above statement [9].

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

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

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

$$\|\hat{x}\| \le \|\hat{e}\|, \quad \hat{x} \in L, \quad \hat{x} \notin \{0, \pm\hat{e}\}, \tag{3}$$

where $\|x\|$ represents Euclidean norm of x. Then $\hat{x}$ satisfies the condition (3) only when $x_{n+1} = 0$, because otherwise we have $\|\hat{x}\| \ge |x_{n+1}| \ge N > \sqrt{n} \ge \|\hat{e}\|$ which contradicts the condition (3). Hence we have some integer y that satisfies

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

Then

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

holds, because

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

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

$$P \le \#\{x \in \mathbb{Z}^n \mid \| x \| \le \| e \|\} \cdot \#\{y \in \mathbb{Z} \mid | y | \le n\sqrt{n/2}\} \times \Pr\left[\sum_{i=1}^{n} a_i z_i = 0\right]. \tag{4}$$

From Lemma 1 in [9], the first term of (4) is bounded by $2^{(\log_2 e)\delta(u)n}$ for any $u \in \mathbb{R}^+$. From Theorem 1 in [9], there exists some value $u_0 \in \mathbb{R}^+$ such that $\delta(u)$ has its minimum value at $u = u_0$. Writing $(\log_2 e)\delta(u_0)$ as c_0, the first term of (4) is bounded by $2^{c_0 n}$. When $z_n = z_{n-1} = \ldots = z_{i+1} = 0$ and $z_i \ne 0$, let z' denote $z' = -\frac{1}{z_i}\sum_{j=1}^{i-1} a_j z_j$, then

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

Hence we can estimate the last term of (4) by

$$\begin{aligned}
&\Pr\left[\sum_{j=1}^{n} a_j z_j = 0\right] \\
&= \sum_{i=1}^{n} \Pr[z_n = z_{n-1} = \ldots = z_{i+1} = 0, z_i \ne 0] \\
&\qquad \times \Pr[a_i = z' \mid z_n = z_{n-1} = \ldots = z_{i+1} = 0, z_i \ne 0] \\
&\le \sum_{i=1}^{n} \Pr[z_n = z_{n-1} = \ldots = z_{i+1} = 0, z_i \ne 0]\frac{1}{A_i} \qquad (5) \\
&\le \sum_{i=1}^{n} \frac{1}{A_i}.
\end{aligned}$$

□

Corollary 1. *Let* $\mathrm{HM}(A_1, \ldots, A_n)$ *denote the harmonic mean of* $A_1, \ldots, A_n$*, i.e.*

$$\mathrm{HM}(A_1, \ldots, A_n) = \frac{1}{\frac{\frac{1}{A_1} + \cdots + \frac{1}{A_n}}{n}}.$$

If for some $c > c_0$*,*

$$\lim_{n\to\infty} \frac{\log_2 \mathrm{HM}(A_1, \ldots, A_n)}{n} = c,$$

then

$$P \to 0 \quad (n \to \infty).$$

Proof. From Theorem 3, we have

$$\lim_{n\to\infty} P \le \lim_{n\to\infty} \frac{n(2n\sqrt{n/2} + 1)2^{c_0 n}}{\mathrm{HM}(A_1, \ldots, A_n)} = 0.$$

□

Above Corollary 1 indicates that in case we choose a_i's from different periods $A_1, \ldots, A_n$, we may use another indicator as its density:

$$d_{\mathrm{HM}} = \frac{n}{\log_2 \mathrm{HM}(A_1, \ldots, A_n)}. \tag{6}$$

If we assume an SVP oracle of lattice, we can asymptotically solve the subset sum problem when d_{HM} is smaller than the critical density $d_0 = 0.6463\ldots$.

The reason why harmonic mean of A_i's appears here is as follows. In the inequality (4) of theorem 3, we bound the third term $\Pr\left[\sum_{i=1}^{n} a_i z_i = 0\right]$ by $\sum_{i=1}^{n} \frac{1}{A_i}$. In theorem 2, the corresponding term is bounded by $\frac{n}{A}$. If we represent

$$\sum_{i=1}^{n} \frac{1}{A_i} = \frac{n}{A'}$$

for some A' and replace A in the definition of usual density (2) with A' in our case, we are able to prove our statement. From above equation, A' coincides with the harmonic mean of A_i's. Hence d_{HM} can be regarded as a natural extension of usual density d, and if all A_i's are the same value A, d_{HM} coincides with d.

Further, we may combine this density with Kunihiro's density [6]

$$D = \frac{n\mathrm{H}(\frac{k}{n})}{\log_2 A},$$

where H is the binary Entropy function $\mathrm{H}(x) = -x\log x - (1-x)\log(1-x)$, and k is the Hamming weight of the solution:

$$D_{\mathrm{HM}} = \frac{n\mathrm{H}(\frac{k}{n})}{\log_2 \mathrm{HM}(A_1, \ldots, A_n)}.$$

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

Theorem 4. *Let $e = (e_1, \ldots, e_n) \neq (0, \ldots, 0) \in \{0,1\}^n$ be fixed. Let $A_1, \ldots, A_n$ be positive integers and $a_1, \ldots, a_n$ 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:*

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

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 $\hat{e}' = (e_1 - \frac{1}{2}, \ldots, e_n - \frac{1}{2}, 0)$ is less than

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

Note that the critical density $d_1 = 0.9408\ldots$ in CJLOSS algorithm case coincides with $\frac{1}{c_1}$ in above statement (Theorem 3.1 in [2]).

Corollary 2. *Let $\mathrm{HM}(A_1, \ldots, A_n)$ denote the harmonic mean of $A_1, \ldots, A_n$. If for some $c > c_1$,*

$$\lim_{n\to\infty}\frac{\log_2 \mathrm{HM}(A_1, \ldots, A_n)}{n} = c,$$

then

$$P \to 0 \quad (n \to \infty).$$

In case of CJLOSS+ algorithm, we use the following function $\delta_\beta(u)$ of $u \in \mathbb{R}^+$ (Theorem 3.1 in [5]):

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

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

In Sect. 2.2, we saw the success probability of CJLOSS+ algorithm when bit lengths of weights are uniformly distributed from 71 to 80. According to definition (1), the density $d \approx \frac{80}{80} = 1$ in this case, but the success probability is smaller than the case where all weights have 80 bit length and the density is almost 1 also. According to our definition of density (6),

$$d_{\mathrm{HM}} \approx \frac{80}{74}$$

in the case weights are uniformly distributed from 71 to 80, and its success probability is closer to the case where all weights have bit length 74 and $d_{\mathrm{HM}} \approx \frac{80}{74}$, than the case all weights have bit length 80 and $d_{\mathrm{HM}} \approx \frac{80}{80} = 1$. This may be rather an ideal case, and in general non-asymptotic case, we need minute examination of the inequality (5) in the proof of Theorem 3. Let P_i denote the probability $\Pr[z_n = z_{n-1} = \ldots = z_{i+1} = 0, z_i \neq 0]$. In the proof, we bounded each term $P_i\frac{1}{A_i}$ with $\frac{1}{A_i}$ in an asymptotic case. However, when we deal with concrete subset sum problems where n is a fixed value, we should rather analyze the coefficients P_i minutely. For example, if the range of the distribution of bit lengths is wide, the value of $\frac{1}{a_i}$ for smaller a_i will get far greater than that of larger a_i, hence the harmonic mean of a_i's will lean to smaller a_i and the effect of larger a_i in the indicator d_{HM} might get smaller than its actual effect.

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.

3.3 Application in the Context of Knapsack Cryptosystems

Possible application of our work in the context of knapsack-type cryptosystems is to use it in the security analysis of the system when we reduce the data size of public keys. If we would like to reduce the total public key size in knapsack-type cryptosystems, we need to have weights with shorter bit lengths. For example, if we have 80 public key weights with 80-bit each, total public key size is 6400 bits. If we have 80 weights with bit lengths between 61 to 80, 4 keys for each bit length, total public key size is 5640 bits, reducing 11.9% of public key data size. In order to assure the security of the system, we need to analyze the hardness of general subset sum problems in our setting.

4. Concluding Remarks

In this paper, we considered the hardness of general subset sum problems against LO algorithm and its variants, with an assumption that weights are chosen from different intervals respectively. In asymptotic case, we introduced another

density that works as an criterion for the success probabilities of LO algorithm and its variants, and obtained some theoretical results. In non-asymptotic case, we saw the effectiveness and concerns of our new density through concrete experiments. Possible application of our result in the context of knapsack cryptosystems is the security analysis when we reduce the data size of public keys.

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 hardness of general subset sum problems more precisely.

References

[1] E.F. Brickell, "Breaking iterated knapsacks," Advances in Cryptology: Proc. 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, vol.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. Inf. Theory, vol.34, no.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] T. Izu, J. Kogure, T. Koshiba, and T. Shimoyama, "Low-density attack revisited," Designs, Codes and Cryptography, vol.43, no.1, pp.47–59, Springer, 2007.
[6] N. Kunihiro, "New Definition of Density on Knapsack Cryptosystems," Progress in Cryptology: Proc. Africacrypt 2008, LNCS 5023, pp.156–173, Springer, 2008.
[7] A.K. Lenstra, H.W. Lenstra, Jr., and L. Lovász, "Factoring polynomials with rational coefficients," Mathematische Annalen, vol.261, pp.515–534, 1982.
[8] J.C. Lagarias, and A.M. Odlyzko, "Solving low-density subset sum problems," J. ACM, vol.32, no.1, pp.229–246, 1985.
[9] J.E. Mazo and A.M. Odlyzko, "Lattice points in high-dimensional spheres," Monatsch. Math., vol.110, pp.47–61, 1990.
[10] T. Okamoto, K. Tanaka, and S. Uchiyama, "Quantum public-key cryptosystems," Advances in Cryptology: Proc. CRYPTO 2000, LNCS 1880, pp.147–165, Springer, 2000.
[11] C.P. Schnorr and M. Euchner, "Lattice basis reduction: Improved practical algorithms and solving subset sum problems," Mathematical Programming, vol.66, pp.181–199, 1994.
[12] C.P. Schnorr and H.H. Hörner, "Attacking the Chor-Rivest cryptosystem by improved lattice reduction," Advances in Cryptology: Proc. EUROCRYPT'95, LNCS 921, pp.1–12, Springer, 1995.
[13] S. Vaudenay, "Cryptanalysis of the Chor — Rivest cryptosystem," J. Cryptology, vol.14, no.2, pp.87–100, 2001.

Jun Kogure received the B.Sc. and M.Sc. degrees in mathematics and the Ph.D. degree in complexity science and engineering from the University of Tokyo in 1985, 1987 and 2012, respectively. He joined Fujitsu Ltd. in 1987 and moved to Fujitsu Laboratories Ltd. in 1998. He was a visiting associate professor and a visiting professor of the University of Tokyo in 2005 and 2007, respectively. He has been a non-full-time lecturer of Chuo University since 2005, and was a non-full-time lecturer of Waseda University from 2007 to 2010. He received Electrical Science and Engineering Promotion Award and IPSJ Kiyasu Special Industrial Achievement Award in 2007. He was a member of Cryptography Research and Evaluation Committees from 2000 to 2007, and has been a secretary of the Committees since 2008. He is a member of IPSJ and JSSAC. His research interests are in cryptography and number theoretic algorithms.

Noboru Kunihiro received his B.E., M.E. and Ph.D. in mathematical engineering and information physics from the University of Tokyo in 1994, 1996 and 2001, respectively. He is an Associate Professor of the University of Tokyo. He was a researcher of NTT Communication Science Laboratories from 1996 to 2002. He was an associate professor of the University of Electro-Communications from 2002 to 2008. His research interest includes cryptography and information security. He received the SCIS'97 Paper Prize and the Best Paper Award of IEICE in 2010.

Hirosuke Yamamoto was born in Wakayama, Japan, on November 15, 1952. He received the B.E. degree from Shizuoka University, in 1975 and the M.E. and Ph.D. degrees from the University of Tokyo, in 1977 and 1980, respectively, all in electrical engineering. In 1980, he joined Tokushima University. He was an Associate Professor at Tokushima University, the University of Electro-Communications, and the University of Tokyo, from 1983 to 1987, from 1987 to 1993, and from 1993 to 1999, respectively. Since 1999, he has been a Professor at the University of Tokyo. He was with the School of Engineering and the School of Information Science and Technology from 1993 to 1999 and from 1999 to 2004, respectively, and is currently with the School of Frontier Sciences in the University of Tokyo. In 1989 and 1990, he was a Visiting Scholar at the Information Systems Laboratory, Stanford University. His research interests are in Shannon theory, data compression algorithms, and cryptology. Dr. Yamamoto is a Fellow of the IEEE. He served as an Associate Editor for Shannon Theory, IEEE Transactions on Information Theory from 2007 to 2010 and the Editor-in-Chief for the IEICE Transactions on Fundamentals from 2010 to 2011. He is currently the Junior President of the Engineering Sciences Society of the IEICE.