Cryptanalysis of the quantum public-key cryptosystem OTU under heuristics from Szemerédi-type statements[*]

Shoichi Kamada[1]

Tokyo Metropolitan University
1-1, Minami-Osawa, Hachioji, Tokyo, 192-0397, Japan
shoichi@tmu.ac.jp

Abstract. The knapsack cryptography is the public-key cryptography whose security depends mainly on the hardness of the subset sum problem. Many of knapsack schemes were able to break by low-density attacks, which are attack methods to use the situation that a shortest vector or a closest vector in a lattice corresponds to a solution of the subset sum problem. For the case when the Hamming weight of a solution for a random instance of the subset sum problem is arbitrary, if the density is less than 0.9408, then the instance can be solvable almost surely by a single call of lattice oracle. This fact was theoretically shown by Coster et al.

In Crypto 2000, Okamoto, Tanaka and Uchiyama introduced the concept of quantum public key cryptosystems and proposed a knapsack cryptosystem, so-called OTU cryptosystem. However, no known algorithm breaks the OTU cryptosystem.

In this paper, we introduced Szemerédi-type assumptions, which are the imitations of the statement of Szemerédi's theorem on arithmetic progressions. From this mathematical point of view, we make clear what the average case and the worst case are. For low density attacks, we give better heuristics for orthogonal lattices than Gaussian heuristics. Consequently, we show that the OTU scheme can be broken under some heuristic assumptions.

Keywords: knapsack cryptography · subset sum problems · low density attacks · additive combinatorics · extremal combinatorics · combinatorial number theory · natural density.

1 Introduction

The knapsack cryptography is the public-key cryptography such as encryption schemes and signature schemes whose security depends mainly on the hardness of the knapsack problems. In the context of knapsack cryptography, the following subset sum problem is mainly used as a knapsack problem. Here let $\mathbb{N} := \{1, 2, \ldots\}$.

[*] Supported by JSPS KAKENHI Grant Number 19J00126.

Definition 1.1 (Subset Sum Problem) *For $(a_1, \ldots, a_s) \subseteq \mathbb{N}^s$ and $C \in \mathbb{Z}$, find $(x_1, \ldots, x_s) \in \{0,1\}^s$ satisfying*

$$x_1 a_1 + \ldots + x_s a_s = C.$$

In the definition of the subset sum problem, if a solution $(x_1, \ldots, x_s) \in \{0,1\}^s$ exists, then we say that C is representable as a subset sum of a (multi)set $\{a_1, \ldots, a_s\}$, that the left hand side is a subset sum of a (multi)set $\{a_1, \ldots, a_s\}$ and that $(x_1, \ldots, x_s)$ is a represetation of C. Notice that the terminology "representation" is sometimes used in knapsack cryptography (for example, [40]) and frequently in mathematics such as combinatorics, number theory and their related areas.

The decisional version of the subset sum problem is NP-complete. In general, NP-hard problems are believed to be not easy using a quantum computer. So, it is very important in post-quantum cryptography.

The knapsack cryptography has been investigated since the proposals of Merkle-Hellman cryptosystems in 1978 [26]. However, many of knapsack cryptosystems are broken because of low density attacks from Lagarias and Odlyzko [22].

Here we mention the notions of densities, shortly. Natively, the density measures denseness of a subset of an ambient set. Let $[N] := \{1, \ldots, N\}$. For $(a_1, \ldots, a_s) \in [N]^s$, let $A = \{a_1, \ldots, a_s\}$ denote an s-element (multi)set. Then we can define the natural density

$$\delta(A) := \frac{s}{N}.$$

Usually, the natural density is defined for a set and not for a multiset. For convenience, we define the natural density also for a multiset. The natural density is often used in mathematics, especially combinatorics, number theory and related research areas. In the context of knapsack cryptography, the following density is often used.

$$d(A) := \frac{s}{\log_2 N}.$$

Notice that we can take $N = \max A$ when A is not random.

We take the following three examples of encryption schemes in knapsack cryptography.

1. The Merkle-Hellman encryption schemes, which are the first knapsack schemes proposed in 1978. The basic Merkle-Hellman scheme was broken by Shamir [39] in 1984.
2. The Chor-Rivest encryption scheme, which was proposed in 1988 and was not broken approximately for a decade. This scheme was partially broken by Schnorr-Hörner [37] and completely by Vaudenay [42, 43].
3. The OTU encryption scheme, which is an example of constructions of a paradigm of quantum public-key cryptosystems. The paradigm and the OTU encryption scheme are introduced by Okamoto, Tanaka and Uchiyama [33].

Table 1. Basic information of the three knapsack encryption schemes

Encryption Scheme and Year	Hamming Weight	Break?	Density $d(A)$
Merkle-Hellman 1978	arbitrary	Yes	about 0.5
Chor-Rivest 1988	fixed	Yes	about $\frac{s}{h \log_2 s}$
OTU 2000	fixed	Yes (the present paper)	at most $\frac{s}{h(\log_2 s - \varepsilon)}$

In the above table, $\varepsilon > 0$ is a real number depending on the parameter setting of the OTU scheme. For more details of the OTU scheme, we discuss Section 5.

It is known that if $d(A)$ is less than some critical bounds, then almost all subset sum instances can be solved by a single call of the lattice oracle. Lagarias and Odlyzko [22] showed that the critical bound is 0.645. Coster et. al. [9] showed that the critical bound is 0.9408.

The insecurity of knapsack schemes is not only the low density case but also the high density case. For example, the subset sum problem can be solved in time complexity $O(s^2N)$ by a dynamic programming ([20]). Hence, knapsack schemes are insecure when N is a polynomial in s.

Moreover, Nguyen and Stern [29] introduced the notion of pseudo densities and showed that the security of knapsack schemes depends on the smallness of the Hamming weight of a solution of the subset sum problem even when the density $d(A)$ is high. Later, Kunihiro [21] gave the compatibility between the density $d(A)$ and the pseudo density, and showed that the weaker lower bound 0.8677 for the density $d(A)$ is necessary for the security of knapsack schemes in the case of any Hamming weight. Consequently, the density must be some moderate value.

It is known that there are exhaustive searches of a solution of the subset sum problem. For example, it is shown in [3] that it can be solved in $\tilde{O}(2^{0.291s})$ bit operations in the classical computer and in [4] that it can be solved in $\tilde{O}(2^{(0.241+o(1))s})$ qubit operations in the quantum computer. Recently in [5], two algorithms are proposed, one is to solve the problem in $\tilde{O}(2^{0.283s})$ bit operations in the classical computer and the other is to solve the problem in $\tilde{O}(2^{0.218s})$ qubit operations in the quantum computer.

Our motivation comes from mathematics, especially, additive combinatorics, extremal combinatorics and combinatorial number theory. In such areas, Szemerédi's theorem on arithmetic progressions [41] is well-known theorem. To consider the hardness of the subset sum problem, we introduce mathematical assumptions which we call Szemerédi-type assumptions. These assumptions are statements which are imitations of the statement of Szemerédi's theorem on arithmetic progressions.

For the subset sum problem, we make clear from Szemerédi-type assumptions what the worst case hardness and the average case hardness are.

The organization of this paper is as follows. In Section 2, we review the low density attacks. In Section 3, we describe the basic experiments when low density attacks fail on the average case or not. In Section 4, we introduce Szemerédi-type assumptions and we give better heuristics for orthogonal lattices than Gaussian heuristics. In Section 5, we show that the OTU scheme can be broken when the heuristics given in Section 4 hold. In Section 6, we describe concluding remarks.

2 Low density attacks

2.1 Basic notions for lattices

A lattice $L \subseteq \mathbb{R}^n$ spanned by s linearly independent vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_s \in \mathbb{R}^n$ is defined as

$$L = \{y_1\boldsymbol{b}_1 + \cdots + y_s\boldsymbol{b}_s \colon y_1, \ldots, y_s \in \mathbb{Z}\},$$

where s is the rank of L and n is the dimension of L. For a vector $\boldsymbol{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$, we call $(\boldsymbol{v}, \boldsymbol{v}) := \sum_{i=1}^n v_i^2$ the norm of $\boldsymbol{v}$, where the Euclidean norm of $\boldsymbol{v}$ is $\|\boldsymbol{v}\| := \sqrt{(\boldsymbol{v}, \boldsymbol{v})}$. The covolume $\mathrm{covol}(L)$ of a lattice L is the volume of a fundamental region of L. $\lambda_1(L)$ is the Euclidean norm of a non-zero shortest vector in L. The Hermite invariant $\gamma(L)$ of a lattice L of rank s is defined by

$$\gamma(L) := \frac{\lambda_1(L)^2}{\mathrm{covol}(L)^{2/s}}.$$

The Hermite constant γ_s is defined by

$$\gamma_s := \max\{\gamma(L) \colon L \text{ is a lattice of rank } s\}.$$

A lattice L of rank s such that $\gamma(L) = \gamma_s$ is called a critical lattice. It is known that for sufficiently large s,

$$\frac{s}{2\pi e} + \frac{\log_e(\pi s)}{2\pi e} + o(1) \leq \gamma_s \leq \frac{1.744s}{2\pi e}(1 + o(1)).$$

Shortest vector problem (SVP) is to find a non-zero shortest vector in a lattice. This problem is known to be NP-hard under randomized reduction [1].

The worst case hardness of SVP is supposed to be the problems for critical lattices and extreme lattices. Related these cases, in mathematics, Voronoi [44] showed that a lattice is extreme if it is perfect and eutactic. By the way, a lattice which is critical or extreme is corresponding to the global maximum or the local maximum, respectively. For more details of lattices, see [27, 8, 24].

2.2 Basic idea of low density attacks

The knapsack cryptography is different from the lattice-based cryptography. It does not require the hardness of lattice problems such as SVP. Lattice problems are regarded as rather easier. This justifies low density attacks.

Let $\mathscr{A} \subseteq [N]^s$. We fix a representation $\boldsymbol{x} := (x_1, \ldots, x_s) \in \{0,1\}^s$. Then the set $\mathscr{A}$ can be regarded as a set of subset sum instances since an instance of the subset sum problem can be defined for a tuple $\boldsymbol{a} := (a_1, \ldots, a_s) \in \mathscr{A}$.

X stands for a kind of low density attacks, namely, LO, CJLOSS and CJLOSS+. where the each of symbols comes from the names of the authors in [22, 9]. For $\boldsymbol{a} = (a_1, \ldots, a_s) \in [N]^s$, put $A = \{a_1, \ldots, a_s\}$. Assume that $C \in \mathbb{Z}$ is representable as a subset sum. Then let $L_{\mathrm{X}}(A; C)$ or $L_{\mathrm{X}}(\boldsymbol{a}; C)$ be a lattice spanned by the rows of the following $(s+1)$-by-$(s+1)$ matrix.

$$\boldsymbol{B}_{\mathrm{X}} = \begin{pmatrix} 1 & 0 & \ldots & 0 & sa_1 \\ 0 & 1 & \ldots & 0 & sa_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & sa_s \\ \beta_{\mathrm{X}} & \beta_{\mathrm{X}} & \ldots & \beta_{\mathrm{X}} & sC \end{pmatrix},$$

with

$$\beta_{\mathrm{X}} = \begin{cases} 0 & \text{if X} = \text{LO}, \\ \dfrac{1}{2} & \text{if X} = \text{CJLOSS}, \\ \dfrac{h}{s} & \text{if X} = \text{CJLOSS+}, \end{cases}$$

where $h = \sum_{i=1}^{s} x_i$ is the Hamming weight of $\boldsymbol{x} \in \{0,1\}^s$. Notice that when X = CJLOSS+, the Hamming weight h must be known.

We define

$$\hat{\boldsymbol{x}} = (x_1 - \beta_{\mathrm{X}}, \ldots, x_s - \beta_{\mathrm{X}}, 0). \tag{2.1}$$

It is clear that $\hat{\boldsymbol{x}} \in L_{\mathrm{X}}(A; C)$. Hence, finding $\hat{\boldsymbol{x}} \in L_{\mathrm{X}}(A; C)$

We say that a low density attack with respect to SVP fails if there exists $\hat{\boldsymbol{y}} \in L_{\mathrm{X}}(A; C)$ satisfying

$$\begin{cases} \hat{\boldsymbol{y}} \notin \{-\hat{\boldsymbol{x}}, \boldsymbol{0}, \hat{\boldsymbol{x}}\}, \\ \|\hat{\boldsymbol{y}}\|^2 \leq \|\hat{\boldsymbol{x}}\|^2. \end{cases} \tag{2.2}$$

Notice that SVP is replaced by uSVP (unique SVP) when C is uniquely representable as a subset sum. Notice that since $\boldsymbol{x} \in \{0,1\}^s$ is fixed,

$$\|\hat{\boldsymbol{x}}\|^2 = \begin{cases} h & \text{if X} = \text{LO}, \\ s/4 & \text{if X} = \text{CJLOSS}, \\ h(1 - h/s) & \text{if X} = \text{CJLOSS} + . \end{cases}$$

Let $\mathscr{A} \subseteq [N]^s$. Then the failure probability $P_{\mathrm{X}}(\mathscr{A}; \boldsymbol{x})$ of a low density attack is defined by

$$P_{\mathrm{X}}(\mathscr{A}; \boldsymbol{x}) = \frac{|\{\boldsymbol{a} \in \mathscr{A} \colon {}^{\exists}\hat{\boldsymbol{y}} \in L_{\mathrm{X}}(\boldsymbol{a}; C) \text{ s.t. } \hat{\boldsymbol{y}} \text{ satisfies (2.2)}\}|}{|\mathscr{A}|}.$$

A low density attack “succeeds almost surely” if

$$P_{\mathrm{X}}(\mathscr{A}, \boldsymbol{x}) \to 0 \quad \text{as } s \to \infty.$$

In other words, the subset sum problem can be solvable almost surely by a single call of a lattice oracle.

Now we explain previous results. Define a combinatorial number $M(s,k)$ by

$$M(s,k) := |\{\boldsymbol{y} \in \mathbb{Z}^s \colon \sum_{i=1}^{s} y_i^2 \leq k\}|.$$

In several contexts, it is shown that

$$P_{\mathrm{X}}(\mathscr{A}; \boldsymbol{x}) = O(s^c \frac{M(s,qs)}{N})$$

for some constants $q > 0$ and $c > 0$, and it is known that

$$M(s,qs) \leq 2^{s/d'}$$

for some constant d' depending on p (for example, see [25]).

The critical case of d' is the bound of $d(A)$ for the low density.

Table 2. the density condition for the “almost sure success” of low density attacks for all Hamming weights $h \leq s/2$ of $\boldsymbol{x} \in \{0,1\}^s$

	$\mathscr{A}$	p	d'
Lagarias and Odlyzko (LO) 1985 [22]	$[N]^s$	1/2	0.645...
Coster et al. (CJLOSS) 1992 [9]	$[N]^s$	1/4	0.9408...
Kogure et al. 2012 [19]	$\prod_{i=1}^{s}[N_i]$	(∗)	

The details of (∗) are as follows. $d(A)$ is replaced by

$$d_{\mathrm{HM}}(A) := \frac{s}{\log_2 \mathrm{HM}(N_1, \ldots, N_s)}$$

where $\mathrm{HM}(N_1, \ldots, N_s)$ is the harmonic mean

$$\mathrm{HM}(N_1, \ldots, N_s) = \frac{s}{\sum_{i=1}^{s} 1/N_i}.$$

Then LO and CJLOSS are applied.

Remark 2.1. Related to X = CJLOSS+, Nguyen and Stern [29] showed that a low density attack “succeeds almost surely” even when the Hamming weight h of $\boldsymbol{x} \in \{0,1\}^s$ is sufficiently small and $d(A)$ is greater than d' in the table. In the above setting, we suppose a single call of a lattice oracle. In [17], a tuple $(\beta_{\mathrm{X}}, \ldots, \beta_{\mathrm{X}})$ in $\boldsymbol{B}_{\mathrm{X}}$ is replaced by a set of polynomially many tuples. In such a setting, we must suppose polynomially many calls of a lattice oracle. However, when a low density attack succeeds almost surely as $s \to \infty$, the optimal value is somewhat smaller but asymptotically 0.9408.

3 The average case hardness of the subset sum problem

In what follows, unless otherwise noted, we identify $A = \{a_1, a_2, \ldots, a_s\}$ with a tuple $(a_1, a_2, \ldots, a_s) \in [N]^s$ satisfying $a_1 \leq a_2 \leq \ldots \leq a_s$.

In this paper, we investigate the following orthogonal lattice.

$$L(A) := \{(y_1, \ldots, y_s) \in \mathbb{Z}^s \colon y_1 a_1 + \cdots + y_s a_s = 0\}.$$

By the mapping $(y_1, \ldots, y_s) \mapsto (y_1, \ldots, y_s, 0)$, we can regard the lattice $L(A)$ as the common sublattice of the lattices $L_{\mathrm{X}}(A; C)$ for all $\mathrm{X} \in \{\mathrm{LO}, \mathrm{CJLOSS}, \mathrm{CJLOSS+}\}$. From this. it is important to investigate basics for low density attacks.

Let $\{A_1, \ldots, A_s\}$ be a partition of $[N]$, i.e. $[N] = \bigcup_{i=1}^{s} A_i$ and $A_i \cap A_j = \emptyset$ $(i \neq j)$. Put $n_i = |A_i|$. Hence, $N = \sum_{i=1}^{s} n_i$. Then maximize the multinomial coefficient

$$\binom{N}{n_1, \ldots, n_s} = \frac{N!}{n_1! \cdots n_s!} \approx 2^{NH(n_1/N, \ldots, n_s/N)},$$

where for $p_i \geq 0$ with $\sum_{i=1}^{s} p_i = 1$,

$$H(p_1, \ldots, p_s) = -\sum_{i=1}^{s} p_i \log_2 p_i.$$

The maximum attains when $n_i = |A_i| \approx N/s$ for all $i = 1, \ldots, s$. So, the case $n_i \approx 1/\delta(A)$ can be regarded as the average case. One of possible choices is $A_i = ((i-1)N/s, iN/s]$ for each i. In this paper, the instances of the subset sum problem for $\mathscr{A} = \prod_{i=1}^{s}((i-1)N/s, iN/s]$ are regarded as the average case.

Now, we present numerical experiments on the minimum norms of orthogonal lattices $L(A)$. We use the following computer resource.

- Computer: hp Mobile Workstation Zbook Studio G3
  - Memory: 16.0GB
  - CPU: Intel(R) Xeon(R) CPU E3-1505M v5 @ 2.80GHz
- Software: SageMath 9.0

To obtain some heuristic law on the minimum norms of orthogonal lattices $L(A)$, we done the following experiments. Let $s = 24 + 8l$ $(l = 0, 1, \ldots, 7)$. The natural density $\delta(A)$ is of the form

$$\delta(A) := \frac{s}{N} = (cs^{\alpha k})^{-1},$$

where $\alpha = 0, 0.8, 0.9, 1.0$, $k = 4, 6, 12$ and c is selected suitably. We do the following procedure 100 times for every s in the case of $\alpha = 0$ and for every (α, s, k) in the case of $\alpha \neq 0$.

- As $A = \{a_1, \ldots, a_s\}$, each a_i chosen uniformly at random from $((i-1)N/s, iN/s]$.
- Finding a shortest (or an approximate shortest) vector $\boldsymbol{v}$ in an orthogonal lattice $L(A)$.

- In the only case of $\alpha \neq 0$, decide whether the (minimum) norm $\boldsymbol{v} \cdot \boldsymbol{v}$ coincides with k or not.

The choices of the values of c are as follows.

- In the case of $\alpha = 0$, we set $c = 1/\delta(A) = 2^{16}, 2^{20}, 2^{24}, 2^{48}$. We find a shortest vector in each lattice according to Algorithm 1.
- In the case of $\alpha \neq 0$, we choose c in the table 3 so that $\log_2 c$ is an integer according to the values in the case of $\alpha = 0$ and $s = 80$. We find a shortest (or an approximate shortest) vector according to Algorithm 2.

Table 3. the values of k and c in the case of $\alpha \neq 0$

α	k	c
0.8	4	$2^{-9} \leq c \leq 2^{-5}$
0.8	6	$2^{-10} \leq c \leq 2^{-6}$
0.8	12	$2^{-17} \leq c \leq 2^{-13}$
0.9	4	$2^{-11} \leq c \leq 2^{-6}$
0.9	6	$2^{-13} \leq c \leq 2^{-8}$
0.9	12	$2^{-24} \leq c \leq 2^{-20}$
1.0	4	$2^{-14} \leq c \leq 2^{-8}$
1.0	6	$2^{-16} \leq c \leq 2^{-11}$
1.0	12	$2^{-31} \leq c \leq 2^{-27}$

Consider the situation that the values of s vary in the 8 ways. In 100 times trials for each of s, we take a look at the heuristic law when the frequency that $\boldsymbol{v} \cdot \boldsymbol{v}$ coincides with k is the maximum. We see that this holds for at least 4 of all the values of s. However, there is no such a heuristic law in the case of $\alpha = 0$. More precisely, for the case of $\alpha = 0$, the value of the minimum norm of $L(A)$ for every s decreases as s increases.

Hence, using linear regressions, we investigate the distributions of $(k, \log_2 c)$ for every $\alpha = 0.8, 0.9, 1.0$. Then we observed that for each $\alpha = 0.8$, 0.9, 1.0, we can see that there exist constants $c_0 > 0$ and $c_1 > 0$ such that

$$1/\delta(A) \approx c s^{k\alpha} \quad (c = 2^{-c_1 k - c_0}). \tag{3.1}$$

In the expression (3.1), we can solve it for k. Then we have the following heuristics.

Heuristic 3.1 (Explicit Version) *Let k be a positive integer. Then $\lambda_1(L(A))^2 = k$ implies that*

$$\lambda_1(L(A))^2 = \frac{\log_2 1/\delta(A) + c_0}{\alpha \log_2 s - c_1},$$

where for some $0.9 < \alpha < 1$, $c_0 > 0$ and $c_1 > 0$ are some constants depending on k and α.

Here we describe Algorithm 1 and Algorithm 2. Let $L'(A)$ be a lattice spanned by the rows of the matrix

$$\boldsymbol{B} := \begin{pmatrix} 1\,0 \dots 0\ sa_1 \\ 0\,1 \dots 0\ sa_2 \\ \vdots\ \vdots\ \ddots\ \vdots\ \ \vdots \\ 0\,0 \dots 1\ sa_s \end{pmatrix}.$$

Notice that $\boldsymbol{y} \in L(A)$ implies $(\boldsymbol{y}, 0) \in L'(A)$, i.e. A lattice $L(A)$ can be regarded as a sublattice of a lattice $L'(A)$. If C is representable as a subset sum, then it is clear that $L'(A)$ is a sublattice of $L_{\mathrm{x}}(A; C)$. Unless otherwise noted, a shortest vector in $L(A)$ is a shortest vector that belongs to $L'(A)$ such that its last component is 0. Here, as a lattice basis reduction algorithm, we use the BKZ algorithm of block size 32 (BKZ-32) in default of SageMath 9.0.

Algorithm 1 Calculation of the minimum norm in the case of $\alpha = 0$.

```
Input: $\boldsymbol{B}$
Output: the value of the (approximate minimum) norm
Apply BKZ-32 to $\boldsymbol{B}$ and let $\boldsymbol{redB}$ be its output.
Let $\hat{\boldsymbol{y}}$ be a shortest vector in $\boldsymbol{redB}$.
norm = $\hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{y}}$.
nwloop = 10 {number of while loops}
while nwloop $\neq 0$ do
    Select a row $\boldsymbol{b}_0$ from $\boldsymbol{redB}$ so that $s$th component of $\boldsymbol{b}_0$ is non-zero.
    Select the $s$th row $\boldsymbol{b}_s$ from $\boldsymbol{B}$.
    Choose a row $\boldsymbol{b}_{i_0}$ randomly from $\boldsymbol{B}$ so that $i_i$th component of $\boldsymbol{b}_0$ is zero.
    $\boldsymbol{c}_1 = \boldsymbol{b}_{i_0} + \boldsymbol{b}_s$, $\boldsymbol{c}_2 = \boldsymbol{b}_{i_0} + \boldsymbol{b}_s$
    Let $\boldsymbol{B}_1$ be a matrix for which $(\boldsymbol{b}_1, \boldsymbol{b}_2)$ in $\boldsymbol{B}$ is replaced by $(\boldsymbol{c}_1, \boldsymbol{c}_2)$.
    Shuffle rows of $\boldsymbol{B}_1$ and let $\boldsymbol{B}_1$ be its output.
    Apply BKZ-32 to $\boldsymbol{B}_1$ and let $\boldsymbol{redB}_1$ be its output.
    Let $\hat{\boldsymbol{y}}$ be a shortest vector in $\boldsymbol{redB}_1$.
    if $\hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{y}} <$ norm then
        norm = $\hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{y}}$
    end if
    nwloop = nwloop $- 1$
end while
return norm
```

To secure knapsack schemes, we require

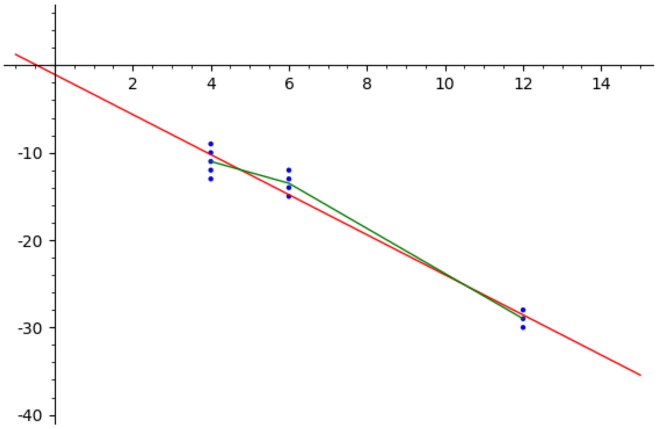
$$d(A) \geq \frac{s}{(h - h^2/s + 1)\log_2 s - (h - h^2/s)c_1 - c_0}, \tag{3.2}$$

but is not necessarily bounded from below by constant 0.9408....

Algorithm 2 Calculation of the minimum norm in the case of $\alpha \neq 0$.

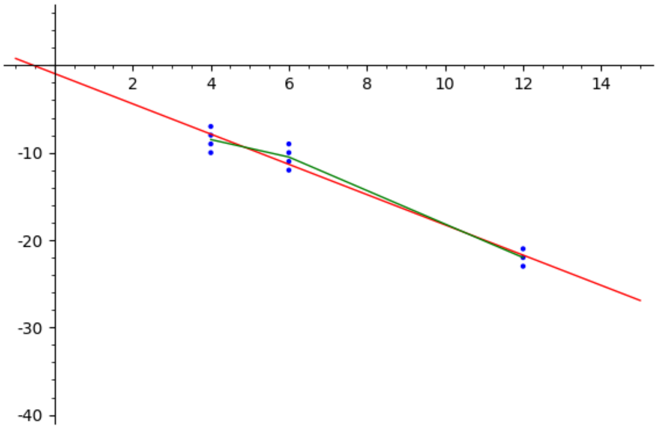
```
Input: $\boldsymbol{B}$, $k$
Output: the value of the (approximate minimum) norm
Apply BKZ-32 to $\boldsymbol{B}$ and let $\boldsymbol{redB}$ be its output.
Let $\hat{\boldsymbol{y}}$ be a shortest vector in $\boldsymbol{redB}$.
norm $= \hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{y}}$.
nwloop $= 10$ {number of while loops}
while nwloop $\neq 0$ do
    if norm $< k$ then
        return norm
    else
        Select a row $\boldsymbol{b}_0$ from $\boldsymbol{redB}$ so that $s$th component of $\boldsymbol{b}_0$ is non-zero.
        Select the $s$th row $\boldsymbol{b}_s$ from $\boldsymbol{B}$.
        Choose a row $\boldsymbol{b}_{i_0}$ randomly from $\boldsymbol{B}$ so that $i_i$th component of $\boldsymbol{b}_0$ is zero.
        $\boldsymbol{c}_1 = \boldsymbol{b}_{i_0} + \boldsymbol{b}_s$, $\boldsymbol{c}_2 = \boldsymbol{b}_{i_0} + \boldsymbol{b}_s$
        Let $\boldsymbol{B}_1$ be a matrix for which $(\boldsymbol{b}_1, \boldsymbol{b}_2)$ in $\boldsymbol{B}$ is replaced by $(\boldsymbol{c}_1, \boldsymbol{c}_2)$.
        Apply BKZ-32 to $\boldsymbol{B}_1$ and let $\boldsymbol{redB}_1$ be its output.
        Let $\hat{\boldsymbol{y}}$ be a shortest vector in $\boldsymbol{redB}_1$.
        if $\hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{y}} <$ norm then
            norm $= \hat{\boldsymbol{y}} \cdot \hat{\boldsymbol{y}}$
        end if
    end if
    nwloop $=$ nwloop $- 1$
end while
if nwloop $= 0$ then
    return norm
end if
```

Fig. 1. $\alpha = 1.0$



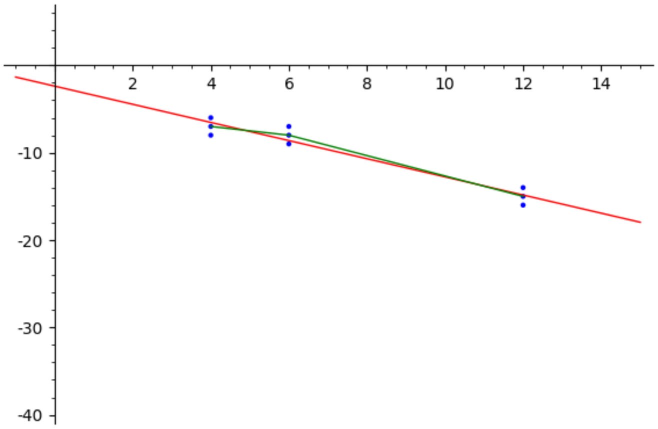
slope $c_1 = -2.2934...$ and y-intercept $c_0 = -1.0434....$

Fig. 2. $\alpha = 0.9$



slope $c_1 = -1.7310...$ and y-intercept $c_0 = -0.9493....$

Fig. 3. $\alpha = 0.8$



slope $c_1 = -1.0384...$ and y-intercept $c_0 = -2.3846....$

4 Mathematical interpretations

4.1 Szemerédi-type assumptions

In this section, we introduce Szemerédi-type assumptions, which are imitations of the statement of Szemerédi's theorem on arithmetic progressions [41].

For a finite set $A = \{a_1, \ldots, a_s\}$, let $\mathbb{Z}^A$ denote the set

$$\{(y_a)_{a\in A} \colon (y_{a_1}, \ldots, y_{a_s}) \in \mathbb{Z}^s\},$$

and put $\mathbf{0} = (0)_{a\in A}$.

For positive integers k and N and a set $A \subseteq [N]$, let $\mathcal{P}(k, N, A)$ be the following statement.

Definition 4.1 (the statement $\mathcal{P}(k, N, A)$) *There exists $(y_a)_{a\in A} \in \mathbb{Z}^A \setminus \{\mathbf{0}\}$ satisfying*

$$\begin{cases} \displaystyle\sum_{a\in A} y_a^2 < k, \\ \displaystyle\sum_{a\in A} y_a a = 0. \end{cases} \tag{4.1}$$

Now we introduce the assumptions which we call Szemerédi-type assumptions. These assumptions come from Szemerédi's theorem on arithmetic progressions. See Appendix A for details of Szemerédi's theorem, and see Appendix B for a kind of the abstract (in)dependence.

Assumption 4.2 (Szemerédi-type assumption of finitary dependence version) *Let δ_0 be a real number with $0 < \delta_0 \leq 1$ and $k \geq 4$ be an integer. Then there exists $N_0(\delta_0, k) > 0$ such that if $N \geq N_0(\delta_0, k)$, then for every set $A \subseteq [N]$ with $|A| \geq \delta_0 N$, the statement $\mathcal{P}(k, N, A)$ is true.*

For $A \subseteq \mathbb{N}$, put $A(N) = A \cap [N]$.

Assumption 4.3 (Szemerédi-type assumption of infinitary (dependence) version) *Let $N_0(\delta_0, k) > 0$ be as above. Then if $A \subseteq \mathbb{N}$ has positive upper asymptotic density, i.e.*

$$\limsup_{N\to\infty} \frac{A(N)}{N} \geq \delta_0$$

for some $0 < \delta_0 \leq 1$, then $N \geq N_0(\delta_0, k)$ implies that the statements $\mathcal{P}(k, N, A(N))$ is true for all $k \geq 4$.

For a positive integer k, let $r(k, N)$ denote the cardinality of a largest set $A \subseteq [N]$ such that the statement $\mathcal{P}(k, N, A)$ is false.

Assumption 4.4 (Szemerédi-type assumption of finitary independence version) *For $k \geq 5$,*

$$r(k, N) = o(N).$$

Notice that $r(4,N)=o(N)$ does not hold since $r(4,N)$ is the cardinality of a largest sum free subset of $[N]$. Indeed, there are two typical examples of maximal sum free sets. One is the set of odd numbers in $[N]$. The other is $(N/2,N]$. Hence, $r(4,N)\geq N/2$, which implies that Szemerédi-type assumption does not hold.

On the other hand, $r(5,N)=o(N)$ holds since any Sidon set in $[N]$ has cardinarity at most $\sqrt{\frac{N}{\log_e N}}$. Hence, $k\geq 5$ is necessary. Moreover, it should hold the equivalence between Assumptions 4.2, 4.3 and 4.4. In [32], there are a lot of information for additive combinatorics and related areas.

Assumption 4.4 will relate solution free sets for a linear equation over positive integers.

Table 4. Examples of solution-free sets for given norms

norm	forbidden equation	structure
2	$v_1=v_2$	a set (of distinct elements)
3	$v_1+v_2=v_3$	a sum free set
4	$v_1+v_2=v_3+v_4$	a Sidon set (B_2 sequence)
6	$v_1+v_3=2v_2$	a progression free set (3-AP free set)

Notice that "variables" $v_1,v_2,\dots$ in the above table mean elements of $[N]$. From this point of view, a set of lattice points in $L(A)$ can be regarded as a set of homogeneous linear equations.

As fundamental combinatorial structures for broadcast encryptions and codes with identifiable parent property there sometimes appear solution free sets to forbid several equations [2, 15] and their improvements are given in [45].

Assume that $A\subseteq[N]$ has cardinality $r(k,N)$ and satisfies property (4.1). Then it is clear that $\lambda_1(L(A))^2\geq k$. We shall regard that $\lambda_1(L(A))^2=k$ since we now interest in high density case. Given the set $A\subseteq[N]$ and a representation $\boldsymbol{x}\in\{0,1\}^A$, $r(k,N)$ can be regarded as a characterization of the worst case hardness of the subset sum problem.

Consider the average case. Assume that an s-element set $A\subseteq[N]$ satisfies $\lambda_1(L(A))^2=k$. Then to guarantee that low density attacks fail, we need

$$4k\leq s\leq r(k,N). \tag{4.2}$$

Next, we state Proposition 4.5. We define the set $A\subseteq[N]$ from the following procedure.

1. We select an element a_1 from the set $[N]$, and then put $A_1=\{a_1\}$.
2. For $i=2,3,\dots$, letting

$$B_i=\left\{-\frac{1}{y_i}\sum_{j=1}^{i-1}y_ja_j : y_j\in\mathbb{Z},\ y_i\neq 0,\ \sum_{j=1}^{i}y_j^2<k\right\},$$

select an element a_i from $[N] \setminus B_i$ inductively when $[N] \setminus B_i \neq \emptyset$, and then put

$$A_i = A_{i-1} \cup \{a_i\}.$$

Here, for some integer t, we assume that a "saturation condition"

$$\begin{aligned} [N] \setminus B_t &\neq \emptyset, \\ [N] \setminus B_{t+1} &= \emptyset \end{aligned}$$

holds, and for this t, put $A = A_t$.

Proposition 4.5 (cf. [7]) *The set $A \subseteq [N]$ above has cardinality t and is a maximal subset of $[N]$ for which there is no vector with norm less than k in its orthogonal lattice $L(A)$. Moreover, it holds that*

$$N \leq \frac{M(t, k-1) - 1}{2}.$$

From Proposition 4.5, we immediately deduce the following corollary.

Corollary 4.6 *If $k - 1 = pt$ for a constant $p > 0$, then there exists $d' > 0$ depending on p such that*

$$t \geq d' \log_2(2N + 1),$$

especially

$$r(k, N) \geq d' \log_2(2N + 1).$$

4.2 The elaboration

Proposition 4.5 can be regarded as the characterization of the worst case hardness of the subset sum problem.

Here we need the exact calculation for $M(s, k)$. Its calculation algorithm is given in [29].

From numerical experiments, we show the relations between N and $M(s, k)$ when each of $k = 4, 6, 12$ coincides with (an approximate value of) $\lambda_1(L(A))^2$ on the average case.

k	minimum of N	maximum of N
4	$2^{-0.6756\ldots} M(s,k)^{0.7331\ldots}$	$2^{-2.3949\ldots} M(s,k)^{1.0124\ldots}$
6	$2^{-1.9556\ldots} M(s,k)^{0.8579\ldots}$	$2^{-2.0464\ldots} M(s,k)^{0.9842\ldots}$
12	$2^{-1.8272\ldots} M(s,k)^{0.9159\ldots}$	$2^{-1.4736\ldots} M(s,k)^{0.9626\ldots}$

Notice that we made sure the above table from the procedure 100 times in Section 3 and linear regressions for $\alpha \neq 0$. There is some possibility that the values of exponential parts are around 0.9408 for sufficiently large s and k. Hence, for the failure of low density attacks, we should suppose that

$$d(A) \gtrapprox 1$$

even if Hamming weight of $\boldsymbol{x} \in \{0,1\}^s$ is arbitrary. So, from our experiments, it may hold that

$$N = \Theta(M(s,k)^{\alpha}) \qquad (\alpha \approx 0.9408)$$

on the average case when $k = \lambda_1(L(A))^2$.

Remark 4.1. Our experiment does not require the value of $C \in \mathbb{Z}$ which is representable as a subset sum of $A \subseteq [N]$. it is numerically shown in several literatures such as [36, 38] that taking into account the value of C, the time consumption is highest. when a Hamming weight h satisfies $h \approx s/2$ and the density $d(A)$ is close to 1. However, our experiments remains some possibility that taking into account the value of C, the time consumption is highest for every fixed Hamming weights $h \leq s/2$.

For more exact analyses, we must make clear the concrete value of α and some hidden constant in an asymptotic notation. For this, we can numerically show that when $\lambda_1(L(A))^2 = qs$ for several constants $0 < q \leq 0.25$,

$$N = \frac{M(s,qs)^{0.9408} - 1}{2},$$

i.e. $\alpha \approx 0.9408$ and a hidden constant is about $1/2$. The following table is its details.

Table 5. Whether $N = \left(M(s,qs)^{0.9408} - 1\right)/2$ holds or not when $\lambda_1(L(A))^2 = qs$.

$q \backslash s$	24	32	40	48	56	64	72	80
0.10	×	○	×	×	○	○	○	×
0.15	○	×	○	○	○	○	○	○
0.20	×	○	○	○	○	○	○	○
0.25	×	○	○	○	○	○	○	○

Notice that we made sure the above table from the procedure 100 times in Section 3 for $\alpha \neq 0$.

Assume $\gcd(A) = 1$. Then we have

$$\sqrt{\frac{s(s-1)(2s-1)}{6\delta(A)}} \leq \operatorname{covol}(L(A)) \leq \sqrt{\frac{s(s+1)(2s+1)}{6\delta(A)}} \tag{4.3}$$

since it holds that

$$\operatorname{covol}(L(A)) = \sqrt{\sum_{i=1}^{s} a_i^2}.$$

Hence, we can expect the following heuristics.

Heuristic 4.7 (Implicit Version (Gaussian Heuristic)) *Let* $0 < q \leq 1/4$. *Then* $\lambda_1(L(A))^2 = qs$ *implies that*

$$\lambda_1(L(A))^2 \sim c\frac{s}{2\pi e}M(s, qs)^{\frac{2\alpha}{s-1}}$$

as $s \to \infty$, *where* $0 < c \leq 1$ *increases as* q *increases.*

Remark 4.2. Heuristic 4.7 is superior to Heuristic 3.1.

5 The security of OTU

In many knapsack encryption schemes, a set $A \subseteq [N]$ is a part of the public key and an integer C representable as a subset sum of A is a ciphertext. The OTU scheme, which has this property, is a knapsack encryption scheme using a number field

5.1 Review of the OTU scheme

In this subsection, we review the OTU scheme for short.

Let K be a number field, let $\mathcal{O}_K$ be its ring of integers and let $\mathfrak{p}$ be a prime ideal in $\mathcal{O}_K$. It is well-known that the residue field $\mathcal{O}_K/\mathfrak{p}$ is isomorphic to the finite field $\mathbb{F}_{p^f}$, where p is a rational prime number below $\mathfrak{p}$ and f is called the residue degree.

In many knapsack schemes, one must make a set $A \subseteq [N]$ as a part of the public key. In the case of the OTU scheme, an s-element set $A \subseteq [N]$ is a set of s distinct discrete logarithms over the residue field $\mathcal{O}_K/\mathfrak{p}$. A more precise description is as follows. Let g be a generator of the multiplicative group $(\mathcal{O}_K/\mathfrak{p})^\times$, let $R(\mathfrak{p}) \subseteq \mathcal{O}_K$ be a complete system of representatives for $\mathcal{O}_K/\mathfrak{p}$, let $\{P_1, \ldots, P_s\}$ be a coprime set in $R(\mathfrak{p})$, i.e. any two elements P_i, P_j $(i \neq j)$ satisfy

$$\gcd(\mathcal{N}(P_i), \mathcal{N}(P_j)) = 1,$$

where $\mathcal{N}(P_i)$ is an ideal norm of the principal ideal (P_i) in $\mathcal{O}_K$. For simplicity, we put $N = p^f$ although the value of N should be $p^f - 1$. Assume that each of elements in $R(\mathfrak{p})$ is written as "a small linear combination of a good integral basis in $\mathcal{O}_K$".

Then by using Shor's algorithm, one must find a set $A = \{a_1, \ldots, a_s\} \subseteq [N]$ such that

$$P_i \equiv g^{a_i} \pmod{\mathfrak{p}}.$$

As described in Table 1, the parameter setting for OTU scheme always follows that

$$d(A) \leq \frac{s}{h(\log_2 s - \varepsilon)} \tag{5.1}$$

for some $\varepsilon > 0$ depending on $P_1, \ldots, P_s$.

Now, we describe (5.1) in details. We are interested in the value of s, which is a common value of the cardinality of a set $A = \{a_1, \ldots, a_s\} \subseteq [N]$ and the cardinality of a coprime set $\{P_1, \ldots, P_s\} \subseteq R(\mathfrak{p})$. Without loss of generality, we consider the following two case.

- each P_i is a prime element above a rational prime number p_i such that
$$\mathcal{N}(P_i) = p_i,$$
up to associate elements in K.
- each P_i is a rational prime number p_i.

Immediately, $K = \mathbb{Q}$ implies the common case in the above.

The first case immediately implies (5.1). All the h-element subset products of $\{P_1, \ldots, P_s\}$ are bounded from above by $\mathcal{N}(\mathfrak{p})$, i.e.

$$\prod_{P \in S} P \leq p^f = \mathcal{N}(\mathfrak{p})$$

for any h-element set $S \subseteq \{P_1, \ldots, P_s\}$.

The second case implies the shortcoming case for $K \neq \mathbb{Q}$ as described in the original paper [33]. In this case, the more precise estimate for the density $d(A)$ is

$$d(A) \leq \frac{s}{hf(\log_2 s - \varepsilon)} \tag{5.2}$$

from the parameter setting of the OTU scheme. Indeed, all the h-element subset products of $\{P_1, \ldots, P_s\}$ are bounded from above by $\mathcal{N}(\mathfrak{p})^{1/f}$, i.e.

$$\prod_{P \in S} P \leq p = \mathcal{N}(\mathfrak{p})^{\frac{1}{f}}$$

for any h-element set $S \subseteq \{P_1, \ldots, P_s\}$ and this implies that

$$\max\{P_1, \ldots, P_s\} \leq 2^{\varepsilon} \mathcal{N}(\mathfrak{p})^{\frac{1}{hf}}$$

for some $\varepsilon > 0$ depending on $P_1, \ldots, P_s$. Hence, we have

$$s \leq 2^{\varepsilon} N^{\frac{1}{hf}}. \tag{5.3}$$

For an arbitrary setting of OTU scheme, (5.3) can be replaced by

$$s \leq 2^{\varepsilon} N^{\frac{1}{h}}. \tag{5.4}$$

From (5.3) and (5.4), we immediately have the lower bound for N, so we obtain (5.2) and (5.1), respectively.

By the way, the original paper [33] shows the only case of $\varepsilon = 0$ as approximate estimates. However, we can expect $\varepsilon \approx 0$. So, we suppose $\varepsilon = 0$.

5.2 Some extremal property on integers in a number field

Here, we describe on coprime sets in $R(\mathfrak{p})$. In [7], the upper and lower estimates for the number of coprime sets in $[N]$ are given. Later, the upper bound is improved in [6]. Let $\pi(N)$ be the number of prime numbers in $[N]$. Then it is easy to see that the number of coprime sets in $[N]$ is at least $2^{\pi(N)}$ since the set of prime numbers not greater than N is one of largest coprime sets and the cardinality is $\pi(N)$. There appears $\pi(N)$ in the estimates due to [7] and [6].

For the secutrity of the OTU, we must consider the number field version of coprime sets. Since s is the cardinality of a coprime set $\{P_1, \ldots, P_s\} \subseteq R(\mathfrak{p})$, it suffices to consider the cardinality for a largest coprime set in $R(\mathfrak{p})$.

Landau showed that the following theorem in [23].

Theorem 5.1 (Landau's prime ideal theorem) *Let K be a number field and let X be a positive real number. Then we have*

$$|\{\mathfrak{p} \colon \mathfrak{p} \textit{ is a prime ideal of } \mathcal{O}_K \textit{ and } \mathcal{N}(\mathfrak{p}) \le X\}| = (1+o(1))\frac{X}{\log_e X}$$

as $X \to \infty$.

To consider the arbitrary settings of the OTU scheme, we are interested in the number of principal ideals in O_K which are prime. From Landau's prime ideal theorem, the number of principal ideals of norm at most X which are prime is at most

$$(1+o(1))\frac{X}{\log_e X}.$$

From (5.4), we put $X = N^{1/h}$. Then it must hold that

$$s \le (1+o(1))\frac{hN^{\frac{1}{h}}}{\log_e N}.$$

Notice that we know Heuristic 4.7. Although $N = p^f$, we put

$$N = \frac{M(s,k)^{\alpha} - 1}{2} \qquad (0.9 \le \alpha \le 1)$$

for simplicity, which includes the average case hardness and the worst case hardness of the subset sum problem. Then we have

$$s \le (1+o(1))\frac{h\left(\frac{M(s,k)^{\alpha}-1}{2}\right)^{\frac{1}{h}} \log_2 e}{\log_2(M(s,k)^{\alpha} - 1) - 1}.$$

Now we assume that $h = qs$ and $k = q(1-q)s$ for some constant $0 < q \le 1/2$. Then simplifying the above estimate, we have

$$s \lessapprox \frac{h\left(M(s,q(1-q)s)\right)^{\frac{\alpha}{h}} \log_2 e}{\alpha \log_2 M(s,q(1-q)s)}.$$

Moreover, we assume that

$$M(s, q(1-q)s) = 2^{\frac{s}{d''}}$$

for some $d'' > 0$ depending on q. Then we have

$$s \lessapprox \frac{qd''2^{\frac{\alpha}{qd''}} \log_2 e}{\alpha}. \tag{5.5}$$

5.3 How to break the OTU scheme

Here we consider the average case of the subset sum problem. Hence, we suppose $\alpha = 0.9408$ since the right hand side of (5.5) takes larger values in the case of smaller values of $\alpha > 0$. Define the function

$$F(x) := \frac{x}{\alpha} 2^{\frac{\alpha}{x}} \log_2 e,$$

where $\alpha = 0.9408$.

From (4.2), we can see that to guarantee the 128bit security, we must have

$$s = \frac{128}{0.218} \frac{1}{H(q)}, \quad k = \frac{128}{0.218} \frac{q(1-q)}{H(q)}.$$

To break OTU scheme, we must show that $s > F(qd'')$. Indeed, such situations hold since it can be easily seen from Figure 4, Figure 5 and Figure 6, where the values of $0 < q \leq 1/2$ are in the horizontal axis in each of the figures. The details for calculations of d'' are as follows. Define the following functions.

$$\theta(z) := 1 + 2\sum_{i=1}^{\infty} z^{i^2}$$

and

$$G(x) := q(1-q)x + \log_e \theta(e^{-x}).$$

Then we can evaluate the value of $d'' > 0$ by

$$d'' := \frac{1}{\min_x G(x)}.$$

By the way, we take a look at critical values for 128 bit security. Then we have

$$s \geq 588, \quad k \leq 147.$$

Fig. 4. the values of s and $F(qd'')$ for the 128bit security of the OTU scheme

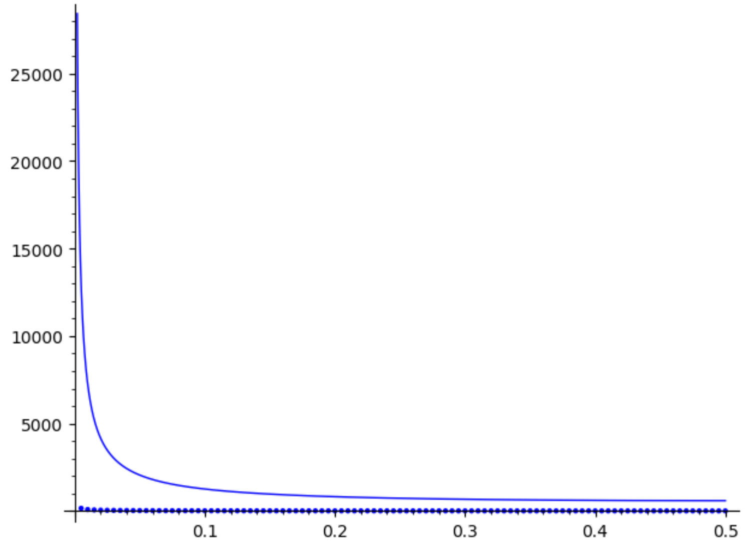


Fig. 5. the values of $F(qd'')$ for the security of the OTU scheme

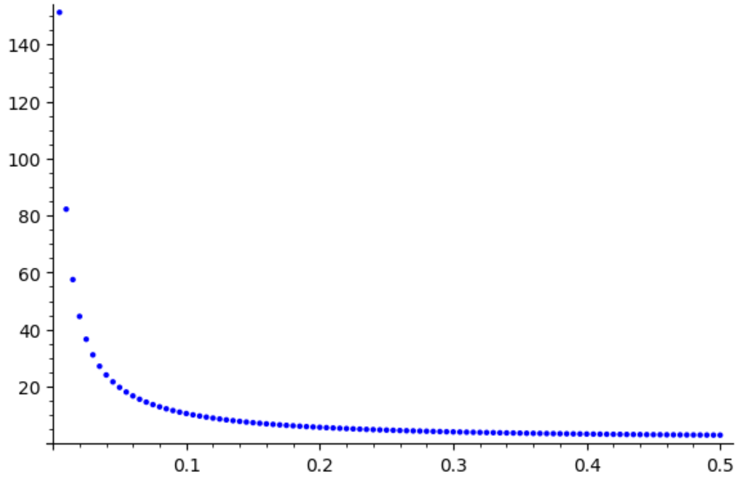
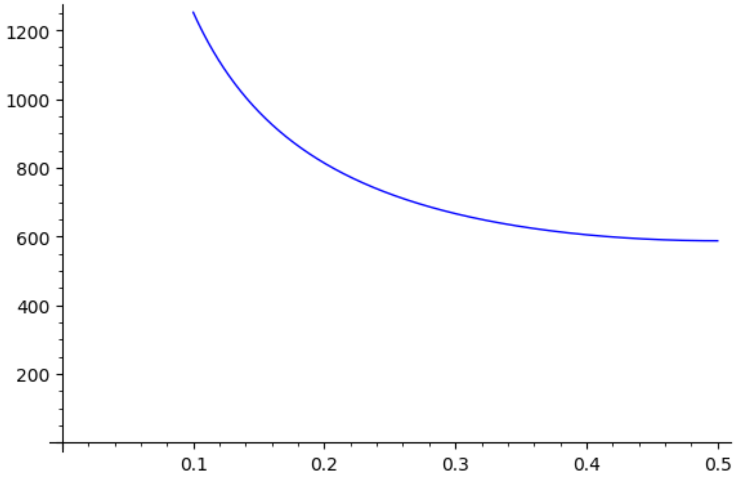


Fig. 6. the values of s for the 128bit security of general knapsack schemes



6 Concluding Remarks

For the OTU scheme, several implementations and improvements were given by the research group of Tokyo Metropolitan University [31, 30, 28]. However, their several implementations do not always work. Compared with them, our heuristic results gave some stronger evidence to them. So, as one of directions, a new format of knapsack schemes may be proposed.

Several knapsack schemes other than the OTU scheme may be broken. For example, in [16], Inoue, the author and Naito proposed a p-adic knapsack encryption scheme. This scheme can be broken.

Here we explain the p-adic knapsack scheme for short. For the detailed description of this scheme, see Appendix C. Here we consider the security of this p-adic knapsack scheme for low density attacks. Let $A = \{a_1, \ldots, a_s\} \subseteq [N]$ be a public key in this scheme. Then the density $d(A)$ can be estimated by

$$d(A) < \frac{s}{m \log_2 p + \log_2 s},$$

where p is a rational prime number and m is a positive integer.

To maximize the density $d(A)$, a rational prime number p must be 2. Hence, we have

$$d(A) < \frac{s}{m + \log_2 s}.$$

Moreover, we must require the smallness of $v_p(b_s)$. In such a case, it is optimal when $s = m$, which implies that $d(A) < 1$, rigorously. One of our main results is that on the average case, $d(A) \gtrapprox 1$ is necessary for the failures of low density attacks. Consequently, the p-adic knapsack scheme is broken.

In this paper, we introduced Szemerédi-type assumptions from which we cryptanalysed the OTU scheme.

The Szemerédi-type assumptions were also to consider the worst case hardness and the average case hardness of the subset sum problem from the aspects of low density attacks. There should be some possibility of breaking a large part of knapsack cryptography. It is difficult for us to mention how many concrete knapsack schemes are broken, except the p-adic knapsack scheme. Moreover, it is not enough to investigate the provable security and the security level (e.g. 128-bit security). We want the reader to find a knapsack scheme which is secure or insecure.

The name of Szemerédi-type assumptions comes from Szemerédi's theorem on arithmetic progressions. For the Szemerédi's theorem, there are many proofs such as combinatorial proofs, Fourier analytic proofs and ergodic theoretic proofs. The first combinatorial proof was given in the Szemerédi's original paper [41]. The Fourier analytic proofs were given in [34, 12]. The first ergodic theoretic proof was given in [10], for which the book [11] is useful for the introductory study.

As a "relative" Szemerédi's theorem, Green-Tao theorem [14] is well known, This theorem states that there exists an arbitrarily long arithmetic progression in the set of prime numbers. For this direction, a large generalization of Green-Tao theorem is appeared in preprint [18], recently. In [18], they focus on the two types of generalizations. One is a generalization of arithmetic progressions. The other is a generalization from $\mathbb{Z}$ to $\mathcal{O}_K$ for a number field K. Since the OTU scheme requires the use of a number field, a new improvement of this scheme may be close to such pure mathematics.

For Proposition 4.5, we imitate the description of a paper due to Cameron and Erdös in 1990 [7], which we refer to CE1990.

CE1990 is also a paper that describes Cameron-Erdös conjecture on sum free subsets of $\mathbb{N} = \{1, 2, \ldots\}$. This conjecture is now a theorem, proved independently by Green [13] and Sapozhenko [35]. The theorem states that the number of sum free subsets of $[N]$ is $O(2^{\frac{1}{2}N})$. Assuming this theorem, the set of sum free subsets of $\mathbb{N}$ has Hausdorff dimension $\frac{1}{2}$.

On the other hand, CE1990 is a paper in pure mathematics (combinatorial number theory), and of course, there is no mention about cryptography. However, many of basic properties in knapsack schemes (Merkle-Hellman, Chor-Rivest, OTU) are described in CE1990. From Szemerédi type assumptions, we need some characterizations for a subset of $2^{\mathbb{N}}$ with Hausdorff dimension 0.

So, we hope some interaction between a large part of mathematics and knapsack cryptography.

Acknowledgements

The author thank Dr. Shohei Satake for the information of references [2, 15, 45] and Mr. Kota Saito for the information of the reference [6]

References

[1] M. Ajtai. "The shortest vector problem in L^2 is NP-hard for randomized reductions". In: *Proceedings of the 30th annual ACM symposium on Theory of computing.* ACM, 1998, pp. 10–19.

[2] N. Alon, E. Fischer, and M. Szegedy. "Parent-Identifying Codes". In: *Journal of Combinatorial Theory, Series A* 95.2 (2001), pp. 349–359.

[3] A. Becker, J. S. Coron, and A. Joux. "Improved Generic Algorithms for Hard Knapsacks". In: *Advances in Cryptology – EUROCRYPT 2011.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 364–385.

[4] D.J. Bernstein, S. Jeffery, T. Lange, and A. Meurer. "Quantum Algorithms for the Subset-Sum Problem". In: *Post-Quantum Cryptography.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 16–33.

[5] X. Bonnetain, R. Bricout, A. Schrottenloher, and Y. Shen. *Improved Classical and Quantum Algorithms for Subset-Sum.* Cryptology ePrint Archive, Report 2020/168. `https://eprint.iacr.org/2020/168`. 2020.

[6] N. J. Calkin and A. Granville. "On the Number of Co-Prime-Free Sets". In: *Number Theory: New York Seminar 1991–1995.* Ed. by David V. Chudnovsky, Gregory V. Chudnovsky, and Melvyn B. Nathanson. New York, NY: Springer US, 1996, pp. 9–18.

[7] P.J. Cameron and P. Erdös. "On the Number of Sets of Integers With Various Properties". In: *Number Theory* (1990), pp. 61 –80.

[8] J.H. Conway and N.J.A. Sloane. *Sphere Packings, Lattices and Groups.* Vol. 290. Grundlehren der mathematischen Wissenschaften. Springer, New York, NY, 1999.

[9] M.J. Coster, A. Joux, B.A. Lamacchia, A.M. Odlyzko, C.P. Schnorr, and J. Stern. "Improved low-density subset sum algorithms". In: *computational complexity* 2 (1992), pp. 111–128.

[10] H. Furstenberg. "Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions". In: *Journal d'Analyse Mathématique* 31 (1977), 204—256.

[11] H. Furstenberg. *Recurrence in Ergodic Theory and Combinatorial Number Theory.* Princeton University Press, 2014.

[12] W. Gowers. "A new proof of Szemerédi's theorem". In: *GAFA, Geometric and functional analysis* 11 (2001), 465—588.

[13] B. Green. "The Cameron–Erdős Conjecture". In: *Bulletin of the London Mathematical Society* 36.6 (Nov. 2004), pp. 769–778.

[14] B. Green and T. Tao. "The Primes Contain Arbitrarily Long Arithmetic Progressions". In: *Annals of Mathematics* 167.2 (2008), pp. 481–547.

[15] Y. Gu and S. Satake. "On 2-parent-identifying set systems of block size 4". In: *Designs, Codes and Cryptography* 88 (2020), 2067—2076.

[16] H. Inoue, S. Kamada, and K. Naito. "Simultaneous approximation problems of p-adic numbers and p-adic knapsack cryptosystems: Alice in p-adic numberland". In: *P-Adic Numbers, Ultrametric Analysis, and Applications* 8 (2016), pp. 312–324.

[17] T. Izu, J. Kogure, T. Koshiba, and T. Shimoyama. “Low-density attack revisited”. In: *Designs, Codes and Cryptography* 43.1 (2007), pp. 47–59.

[18] W. Kai, M. Mimura, A. Munemasa, S. Seki, and K. Yoshino. *Constellations in prime elements of number fields.* 2020. arXiv: `2012.15669 [math.NT]`.

[19] J. Kogure, N. Kunihiro, and H. Yamamoto. “On the hardness of subset sum problem from different intervals”. In: *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences* E95-A.5 (2012), pp. 903–908.

[20] B. Korte and J. Vygen. *Combinatorial Optimization: Theory and Algorithms.* 6th ed. Springer-Verlag Berlin Heidelberg, 2018.

[21] N. Kunihiro. “New Conditions for Secure Knapsack Schemes against Lattice Attack”. In: *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences* E93-A.6 (2010), pp. 1058–1065.

[22] J.C. Lagarias and A.M. Odlyzko. “Solving low-density subset sum problems”. In: *Journal of the ACM* 32 (1985), pp. 229–246.

[23] E. Landau. “Neuer Beweis des Primzahlsatzes und Beweis des Primidealsatzes”. Writen in German. In: *Mathematische Annalen* 56 (1903), pp. 645–670.

[24] J. Martinet. *Perfect Lattices in Euclidean Spaces.* Vol. 327. A Series of Comprehensive Studies in Mathematics. Springer, Berlin, Heidelberg, 2003.

[25] J.E. Mazo and A.M. Odlyzko. “Lattice points in high-dimensional spheres”. In: *Monatshefte für Mathematik* 110 (1990), pp. 47–61.

[26] R.C. Merkle and M.E. Hellman. “Hiding information and signatures in trapdoor knapsacks”. In: *IEEE Transactions on Information Theory* 24 (1978), pp. 525–530.

[27] J. Milnor and D. Husemoller. *Symmetric Bilinear Forms.* Vol. 73. Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer, Berlin, Heidelberg, 1973.

[28] Y. Miyamoto and K. Nakamula. “Improvement of key generation for a number field based knapsack cryptosystem”. In: *JSIAM Letters* 5 (2013), pp. 45–48.

[29] P.Q. Nguyen and J. Stern. “Adapting Density Attacks to Low-Weight Knapsacks”. In: *Advances in Cryptology – ASIACRYPT 2005.* Springer Berlin Heidelberg, 2005, pp. 41–58.

[30] K. Nishimoto and K. Nakamula. “On a knapsack based cryptosystem using real quadratic and cubic fields”. In: *JSIAM Letters* 2 (2010), pp. 81–84.

[31] K. Nishimoto and K. Nakamula. “On key generation of OTU2000 and related problems”. Writen in Japanese. In: *Transactions of the Japan Society for Industrial and Applied Mathematics* 18.1 (2008), pp. 185–197.

[32] K. O’Bryant. “A Complete Annotated Bibliography of Work Related to Sidon Sequences”. In: *The Electronic Journal of Combinatorics* DS11 (Dynamic Surveys 2004).

[33] T. Okamoto, K. Tanaka, and S. Uchiyama. “Quantum Public-Key Cryptosystems”. In: *Advances in Cryptology - CRYPTO 2000.* Springer Berlin Heidelberg, 2000, pp. 147–165.

[34] K.F. Roth. “On Certain Sets of Integers”. In: *Journal of the London Mathematical Society* s1-28.1 (1953), pp. 104–109.

[35] A.A. Sapozhenko. “The Cameron–Erdős conjecture”. In: *Discrete Mathematics* 308.19 (2008). Simonovits ’06, pp. 4361–4369.

[36] C.P. Schnorr and M. Euchner. “Lattice basis reduction: Improved practical algorithms and solving subset sum problems”. In: *Mathematical Programming* 66 (1994), pp. 181–199.

[37] C.P. Schnorr and H.H. Hörner. “Attacking the Chor-Rivest Cryptosystem by Improved Lattice Reduction”. In: *Advances in Cryptology — EUROCRYPT ’95.* Springer Berlin Heidelberg, 1995, pp. 1–12.

[38] C.P. Schnorr and T. Shevchenko. “Solving Subset Sum Problems of Density close to 1 by “randomized” BKZ-reduction”. In: *IACR Cryptology ePrint Archive* 2012 (2012), p. 620.

[39] A. Shamir. “A Polynomial-Time Algorithm for Breaking the Basic Merkle-Hellman Cryptosystem”. In: *IEEE Transactions on Information Theory* 30 (1984), pp. 699–704.

[40] A. Shamir. “On the Cryptocomplexity of Knapsack Systems”. In: *Proceedings of the 11th Annual ACM Symposium on the Theory of Computing.* ACM, 1979, pp. 118–129.

[41] E. Szemerédi. “On sets of integers containing no k elements in arithmetic progression”. In: *Acta Arithmetica* 27 (1975), pp. 199–245.

[42] S. Vaudenay. “Cryptanalysis of the Chor-Rivest cryptosystem”. In: *Advances in Cryptology — CRYPTO ’98.* Springer Berlin Heidelberg, 1998, pp. 243–256.

[43] S. Vaudenay. “Cryptanalysis of the Chor-Rivest Cryptosystem”. In: *Journal of Cryptology* 14 (2001), pp. 87–100.

[44] G. Voronoi. “Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Premier mémoire. Sur quelques propriétés des formes quadratiques positives parfaites.” Writen in French. In: *Journal für die reine und angewandte Mathematik* 1908.133 (1908), pp. 97–102.

[45] X. Wang. “Improved upper bounds for parent-identifying set systems and separable codes”. In: *Designs, Codes and Cryptography* 89.1 (2021), pp. 91–104.

A Szemerédi’s theorem on arithmetic progressions

Here we describe Szemerédi’s theorem on arithmetic progressions [41]. The following Theorem A.1, Theorem A.2 and Theorem A.3 are equivalent.

Theorem A.1 (Szemerédi’s theorem of finitary dependence version) *Let δ_0 be a real number with $0 < \delta_0 \leq 1$ and $k \geq 3$ be an integer. Then there exists $N_0(\delta_0, k)$ such that if $N \geq N_0(\delta_0, k)$, then for every set $A \subseteq [N]$ with $|A| \geq \delta_0 N$, the set A contains an arithmetic progression of length k.*

For a positive integer k, let $r(k, N)$ denote the cardinality of a largest set $A \subseteq [N]$ such that the set A does not contain an arithmetic progression of length k.

Theorem A.2 (Szemerédi's theorem of finitary independence version) *For $k \geq 3$,*

$$r(k, N) = o(N).$$

Theorem A.3 (Szemerédi's theorem of infinitary (dependence) version) *If $A \subseteq \mathbb{N}$ has positive upper asymptotic density, i.e.*

$$\limsup_{N \to \infty} \frac{A(N)}{N} > 0,$$

then the set A contains arbitrarily long arithmetic progressions, i.e. the set A contains arithmetic progressions of length k for all $k \geq 3$.

B The notion of independence systems

For a finite X, let $\mathcal{I} \subseteq 2^X$. The definition of the independence system is as follows.

Definition B.1 (Independence System) *An ordered pair $(X, \mathcal{I})$ is an independence system if the following conditions hold.*

1. *$\emptyset \in \mathcal{I}$.*
2. *If $I_2 \in \mathcal{I}$ and $I_1 \subseteq I_2$, then $I_1 \in \mathcal{I}$.*

Given an independence system $(X, \mathcal{I})$, each set $I \in \mathcal{I}$ is called an independent set and each set $D \in 2^X \setminus \mathcal{I}$ is called a dependent set.

C The p-adic knapsack scheme

In the original paper [16], the p-adic knapsack scheme requires a Diophantine inequality for a linear form over the ring of p-adic integers. Here, although we do not require the advanced knowledge of the general theory of Diophantine approximations, we review p-adic knapsack scheme for short.

Let p be a rational prime number. For $x \in \mathbb{N}$, let $v_p(x)$ be the p-adic (additive) valuation of x, i.e. the largest value of n such that

$$x = p^n x', \quad p \nmid x'.$$

Notice that p-adic absolute value of $x \in \mathbb{N}$ is $|x|_p := p^{-v_p(x)}$.

Now we describe the p-adic knapsack scheme. For a positive integer m, let $\{b_1, \ldots, b_s\} \subseteq [p^m - 1]$ be an increasing sequence in the following sense.

$$v_p(b_1) < v_p(b_2) < \cdots < v_p(b_s),$$

which we call p-adic increasing sequence with respect to $v_p(\cdot)$.

For a large rational prime $N > sp^m$, let r be a random integer such that $\gcd(p, r) = 1$. Let t be the inverse of r modulo N, i.e. $tr \equiv 1 \pmod{N}$.

The sequence $\{a_1, \ldots, a_s\} \subseteq [N-1]$ is calculated by

$$a_i \equiv rb_i \pmod{N}.$$

The private key is the tuple $(p, q, m, t, b_1, \ldots, b_s)$. The public key is the tuple $(a_1, \ldots, a_s)$. For the encryption, the ciphertext $C \in \mathbb{Z}$ of the p-adic scheme is given by

$$C = x_1a_1 + \cdots + x_sa_s.$$

For the decryption, "the v_p version of Merkle-Hellman type decryption" is required.