Density Attack on the Knapsack Cryptosystems with Enumerative Source Encoding (Extended Abstract)

Keiji Oomura and Keisuke Tanaka *

Dept. of Mathematical and Computing Sciences
Tokyo Institute of Technology
2-12-1 Ookayama, Meguro-ku, Tokyo 152-8552, Japan
{oomura8,keisuke}@is.titech.ac.jp

Feburary 14, 2003

1 Introduction

Knapsack cryptosystems are based on the subset-sum problem, that is, given a set of integers $a_1, \dots, a_n$ and a specified sum s, find a subset of $\{a_1, \dots, a_n\}$ that sums to exactly s, or equivalently find a 0-1 vector $(e_1, \dots, e_n)$ such that $\sum_{i=1}^{n} a_i e_i = s$.

Merkle and Hellman discovered a way to use the subset-sum problem as the basis for a public key cryptography [7]. Although the underlying problem is NP-complete, it was broken by Shamir [10]. Later, many other variants have been proposed and shown insecure for any practical parameters by lattice reduction techniques. Actually, subset-sum problems can be characterized by the density parameter defined as the ratio of the number of elements in it to the size in bits of these elements.

For solving knapsacks of low-density, two algorithms have been originally proposed, one by Brickell [1] and the other by Lagarias and Odlyzko [5]. Those algorithms can solve almost all low-density subset-sum problems. Both the Brickell and the Lagarias–Odlyzko algorithms reduce the subset-sum problem to that of finding a short vector in a lattice. For finding a short vector, we use the algorithm of Lenstra, Lenstra, and Lovász [6] because it is currently the algorithm that has been rigorously proved both to have a polynomial running time and to find reasonably short vectors in a lattice. Consider a lattice oracle that, given a basis for a lattice, with high probability yields in polynomial-time the shortest non-zero vector in the lattice. The analysis of [5] showed that availability of such an oracle would let the Lagarias–Odlyzko algorithm solve almost all subset-sum problems of density$< 0.6463\cdots$, but not higher than that. (A similar analysis is not available for the Brickell algorithm [1], although it seems to require even lower densities.) Moreover, Coster, Joux, LaMacchia, Odlyzko, Schnorr, and Stern analyzed a simple modification of the Lagarias–Odlyzko algorithm that reduces the subset-sum problem to that of finding a short vector in a lattice problem [3]. They showed that a single call to a lattice oracle would lead to polynomial-time solutions of almost all problems of density $< 0.9408\cdots$.

*Supported in part by Grant-in-Aid for Scientific Research, Ministry of Education, Culture, Sports, Science and Technology, 14780190.

In contrast to the previous cryptosystems, the Chor–Rivest cryptosystem [2] is an example which achieves a density close to 1. Its underlying problem has however the restriction that the subsets must have cardinality equal to h. Refinement of lattice reduction tools with this restriction have been studied by Schnorr and Hörner [9]. They showed that implementation of the Chor–Rivest cryptosystem with some parameters could be broken within a few days of computation on a single workstation (1995). Vaudenay showed how to break the Chor–Rivest cryptosystem with all the originally suggested parameters by using the property of symmetries in the secret keys and by the partial key disclosure attack [11].

However, the Chor–Rivest cryptosystem with reasonably large parameters is considered to be still secure against the low-density attack. Recently, Okamoto, Tanaka, and Uchiyama proposed a knapsack cryptosystem which also achieves a density close to 1. The attack against their system has not been found, and is also considered to be secure against the low-density attack. Notice that both of the Chor–Rivest and the Okamoto–Tanaka–Uchiyama schemes use the enumerative source encoding which transforms a plain text into the 0-1 vectors with length p and weight h ($< p$) [4].

In this paper, we analysis the Lagarias–Odlyzko low-density attack precisely, and show that this low-density attack can be applied to the Chor–Rivest and the Okamoto–Tanaka–Uchiyama cryptoschemes, which are considered to be secure against the low-density attack. According to our analysis, these schemes turn out to be no longer secure against the Lagarias–Odlyzko low-density attack.

2 Low-Density Attack

In this paper, we are concerned with the low-density attacks, which is related to the subset-sum problem and the 0-1 integer programming problem.

Subset Sum Problem

Given: $A = \{a_i \in Z : 1 \leq i \leq n\}$, $M \in Z$.

Question: Is the sum of the elements in some subset of A equal to M?

0-1 Integer Programming Problem

Given: $A = \{a_i \in Z : 1 \leq i \leq n\}$, $M \in Z$.

Find: x such that $\sum_{i=1}^{n} a_i x_i = M$, $x_i = \{0, 1\}$.

The subset sum problem is to decide whether or not the 0-1 integer programming problem has a solution. This problem is NP-complete, and the difficulty of solving it is the basis of public-key cryptosystems of knapsack type.

The low-density attack converts this problem to one of finding a particular short vector v in a lattice, and then to attempt to find v we uses a lattice basis reduction algorithm due to Lenstra, Lenstra, and Lovász [6].

Definition 1 *A lattice* $L \subset \boldsymbol{R}^n$ *such that*

$$L = \left\{ \sum_{i=1}^{n} x_i b_i \middle| x_i \in \boldsymbol{Z}, i = 1, \ldots, n \right\},$$

$b_1, \ldots, b_n \in \boldsymbol{R}^n$ *is linearly independent,* $B = (b_1, \ldots, b_n) \subset \boldsymbol{R}^{n \times n}$ *is the basis of* $L = L(B)$*.*

We briefly review two algorithms which would solve the subset-sum problem with sufficiently low density in polynomial-time by finding a shortest non-zero vector in a lattice. One is the Lagarias–Odlyzko algorithm and the other is the Coster–Joux–LaMacchia–Odlyzko–Schnorr–Stern one.

2.1 Lagarias–Odlyzko Algorithm

Subset-sum problems can be characterized by the density parameter defined as the ratio of the number of elements in it to the size in bits of these elements.

Definition 2 *The density of weights* $a_1, \ldots, a_n$ *is defined by* $d = \frac{n}{\log_2 \max_i a_i}$.

Lagarias and Odlyzko showed that if the density is bounded by $0.6463\ldots$, the lattice oracle is guaranteed to find the solution vector with high probability.

Theorem 3 (Lagarias–Odlyzko[5]) *Let* A *be a positive integer, and let* $a_1, \ldots, a_n$ *be random integers with* $0 < a_i \leq A$ *for* $1 \leq i \leq n$. *Let* $\boldsymbol{e} = (e_1, \ldots, e_n) \in \{0,1\}^n$ *be arbitrary, and let* $s = \sum_{i=1}^n e_i a_i$. *If the density* $d < 0.6463\cdots$, *then the subset-sum problem defined by* $a_1, \ldots, a_n$ *and* s *can 'almost always' be solved in polynomial-time with a single call to a lattice oracle.*

We define $S_n(R)$ to be the number of integer solutions to the inequality

$$\sum_{i=1}^{n} x_i^2 \leq R,$$

that is, the number of integer lattice points inside or on the n-dimensional sphere of radius $\sqrt{R}$ centered at the origin.

Theorem 4 (Lagarias–Odlyzko[5]) *If* $c = \min_{u \in R}(\log_2 e)\delta(\beta, u),\ \delta(\beta, u) = u\beta + \ln\theta(e^{-u}),\ \text{and}\theta(z) = 1 + 2\sum_{k=1}^{\infty} z^{k^2}$, *then for all* $n \geq 1$, $S_n(\beta n) \leq 2^{cn}$.

Proof. From the equation $S_n(\beta n) \leq e^{n\delta} = 2^{(\log_2 e)\delta(\beta,u)n}$ (Equation 3.36 in the paper by Lagarias and Odlyzko [5]). $\diamondsuit$

The low-density attack of Lagarias and Odlyzko is as follows.

Lagarias–Odlyzko algorithm$(a_1, \ldots, a_n, s)$

1. Input: $a_1, \ldots, a_n, s$; Output: $e_1, \ldots, e_n$.
2. Choose $N > \sqrt{n}$.
3. Make the lattice with the following vectors:

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

4. Using the LLL algorithm, find a shortest non-zero vector $v = (v_1, \ldots, v_{n+1})$. If $s = \sum_{i=1}^n a_i v_i$, it will become $v = e$.

2.2 Coster–Joux–LaMacchia–Odlyzko–Schnorr–Stern Algorithm

Coster, Joux, LaMacchia, Odlyzko, Schnorr, and Stern proposed a simple modification of the Lagarias–Odlyzko algorithm to show that almost all subset sum problem of density $< 0.9408\ldots$ can be solved.

Theorem 5 [3] *Let A be a positive integer, and let $a_1, \ldots, a_n$ be random integers with $0 < a_i \leq A$ for $1 \leq i \leq n$. Let $\boldsymbol{e} = (e_1, \ldots, e_n) \in \{0,1\}^n$ be arbitrary, and let $s = \sum_{i=1}^{n} e_i a_i$. If the density $d < 0.9408\cdots$, then the subset-sum problem defined by $a_1, \ldots, a_n$ and s can 'almost always' be solved in polynomial-time with a single call to a lattice oracle.*

This improved attack uses the vector $b_{n+1} = (\frac{1}{2}, \ldots, \frac{1}{2}, Ns)$ instead of the vector b_{n+1} of the Lagarias–Odlyzko algorithm. In almost all subset-sum problems of density $d < 0.9408\cdots$, the solution vector $v = e$ we searched for is a shortest nonzero vector in the lattice. One way to improve the bound presented above would be to show that it is possible to cover the vertices of the n-cube with a polynomial number of the n-spheres of radius $\sqrt{\alpha n}$ with $\alpha < \frac{1}{4}$. But, according to the following proposition, any n-sphere of radius $\sqrt{\alpha n}$ with $\alpha < \frac{1}{4}$ can cover only an exponentially small fraction of the vertices of the n-cube. They claim that it is impossible to improve the bound of density [3].

Proposition 6 [3] *Any sphere of radius $\sqrt{\alpha n}, \alpha < \frac{1}{4}$, in R^n contains at most $(2-\delta)^n$ points of $\{0,1\}^n$, for some $\delta = \delta(\frac{1}{4}, \alpha) > 0$.*

However, if we can restrict the solution vectors to be with weight h ($\leq \frac{n}{4}$), it might be possible to improve the bound of density. This motivates our research. In the Chor–Rivest [2] and the Okamoto–Tanaka–Uchiyama [8] cryptosystems, in order to achieve a density close to 1, the weight h must be less than $\frac{n}{4}$.

3 Precise Analysis of the Lagarias–Odlyzko Low-Density Attack

In the cases where the subset-sum problem to be solved is known to have $\sum_{i=1}^{n} e_i$ small (as occurs in some knapsack cryptosystems, such as the Chor–Rivest [2] and the Okamoto–Tanaka–Uchiyama [8] ones), it is possible to improve on the result of the Lagarias–Odlyzko algorithm [5]. If $\sum_{i=1}^{n} e_i \leq \beta n$ for $0 < \beta \leq \frac{1}{4}$ in the Chor–Rivest and the Okamoto–Tanaka–Uchiyama cryptosystems, we can solve the subset-sum problem of density above $0.9408\ldots$ using the Lagarias–Odlyzko algorithm. In these cryptosystems, in order to achieve a density close to 1, parameter β must be less than $\frac{1}{4}$. To show the main theorem, we need the following observation.

Observation 7 *One n-sphere with a radius $\sqrt{\beta n}$ centered at $c = (0, 0, \ldots, 0)$ can cover the points of $e \in \{0,1\}^n$ for $\sum_{i=1}^{n} e_i \leq \beta n$.*

Proof. The distance h between $c = (0, 0, \ldots, 0)$ and the points $e = (e_1, \ldots, e_n)$ with $\sum_{i=1}^{n} e_i \leq \beta n$ is

$$\begin{aligned} h &= \|e - c\| \\ &= \|e\| \\ &\leq \sqrt{\beta n}. \end{aligned}$$

$\diamondsuit$

Now, we prove the main theorem.

Theorem 5 *Let A be a positive integer, and let $a_1, \dots, a_n$ be random integers with $0 < a_i \le A$ for $1 \le i \le n$. Let $\boldsymbol{e} = (e_1, \dots, e_n) \in \{0,1\}^n$ such that $\sum_{i=1}^n e_i \le \beta n$ for $0 < \beta \le \frac{1}{4}$, and let $s = \sum_{i=1}^n e_i a_i$. If density $d < d_a$ described below, then the subset sum problem defined by $a_1, \dots, a_n$ and s can 'almost always' be solved in polynomial time with a single call to a lattice oracle:*

$$d_a = \max_{u \in R} \frac{1}{(\log_2 e)\delta(\beta, u)}, \ \ \delta(\beta, u) = u\beta + \ln \theta(e^{-u}), \ \ \theta(z) = 1 + 2\sum_{k=1}^{\infty} z^{k^2}.$$

Proof. We can prove this theorem by a similar argument in the proof of Theorem 3. Notice that the solution vector $\hat{e} = (e_1, \dots, e_n, 0)$ is in L. We are interested in vectors $\hat{x} = (x_1, \dots, x_{n+1})$ which satisfy:

$$\begin{cases} \|\hat{x}\| \le \|\hat{e}\|, \\ \hat{x} \in L, \\ \hat{x} \notin \{0, \hat{e}, -\hat{e}\}. \end{cases} \tag{1}$$

We have $\sum_{i=1}^n e_i \le \beta n$, so $\|\hat{e}\| \le \sqrt{\beta n}$. We show that probability P that a lattice L contains a short vector which satisfies Equation 1 is

$$\begin{aligned} P = &\Pr(\exists \hat{x} \textit{ which satisfies Equation 1}) \\ \le &\Pr(\exists \hat{x}, y \textit{ s.t. } \|\hat{x}\| \le \|\hat{e}\|, |y| \le n\sqrt{\beta n}, \hat{x} \notin \{0, \hat{e}, -\hat{e}\}, \sum_{i=1}^n x_i a_i = ys) \\ \le &\Pr\left(\sum_{i=1}^n x_i a_i = ys \middle| 0 < \|\hat{x}\| \le \|\hat{e}\|, |y| \le n\sqrt{\beta n}, \hat{x} \notin \{0, \hat{e}, -\hat{e}\} \right) \\ &\cdot \left| \{\hat{x} \mid \|\hat{x}\| \le \|\hat{e}\| \ \le \sqrt{\beta n}\} \right| \cdot \left| \left\{ y \middle| |y| \le n\sqrt{\beta n} \right\} \right| \\ \le &n\Big(2n\sqrt{\beta n} + 1\Big)\frac{1}{A} \cdot \left| \{\hat{x} \mid \|\hat{x}\| \le \|\hat{e}\| \ \le \sqrt{\beta n}\} \right|. \end{aligned}$$

Let $x = (x_1, \dots, x_n)$ denote an element of Z^n. (Note that if $\hat{x} = (x_1, \dots, x_n, 0)$, then $\|\hat{x}\| = \|x\|$ and as a special case we have $\|\hat{e}\| = \|e\|$.) From Theorem 4, we have $\left| \{\hat{x} \mid \|\hat{x}\| \le \|\hat{e}\| \ \le \sqrt{\beta n}\} \right| = \left| \{x \mid \|x\| \le \|e\| \ \le \sqrt{\beta n}\} \right| \le 2^{c_1 n}$ such that

$$c_1 = \min_{u \in R} (\log_2 e) \Big\{ u\beta + \ln \Big\{ 1 + 2\sum_{k=1}^{\infty} (e^{-u})^{k^2} \Big\} \Big\}.$$

Then,

$$P \le n\Big(2n\sqrt{\beta n} + 1\Big)\frac{2^{c_1 n}}{A}.$$

This implies that, if $A = 2^{c_2 n}$ with $c_2 > c_1$, we have $\lim_{n \to \infty} P = 0$, and the density d is as follows:

$$d = \frac{n}{\log_2 \max\limits_{1 \le i \le n} a_i} = \frac{n}{\log_2 A} < \frac{n}{\log_2 2^{c_1 n}} = \frac{1}{c_1} = d_a.$$

♢

The density d_k of the Chor–Rivest and the Okamoto–Tanaka–Uchiyama[1] cryptosystems depends on n and β such that $\sum_{i=1}^{n} e_i = \beta n$ for encoded text $e = (e_1, \ldots, e_n)$. That is,

$$d_k = \frac{n}{\beta n \log n} = \frac{1}{\beta \log n}.$$

Recall that, from theorem 8,

$$d_a = \max_{u \in R} \frac{1}{(\log_2 e)\delta(\beta, u)}.$$

We show that, for any reasonable parameters n and β, $d_a - d_k > 0$, and the attack of the knapsack schemes can be succeeded. Let $d'_a(\beta, u) = \frac{1}{(\log_2 e)\delta(\beta,u)}$. We analyze the function $z(\beta, u) = d'_a - d_k$ by drawing the graphs. These graphs show that, in reasonable range of β (the horizontal axis), $z(\beta, u) = d'_a - d_k > 0$. For example, if $n = 64$, $\beta = \frac{1}{6}$, Figure 1 shows that the subset-sum problem of density $d = 1$ becomes insecure against the density attack since $d_a - d_k > 0$. We also draw the graphs in the case of $n = 128, 256, 512, 1024, 2048$ in Figure 2–6, respectively.

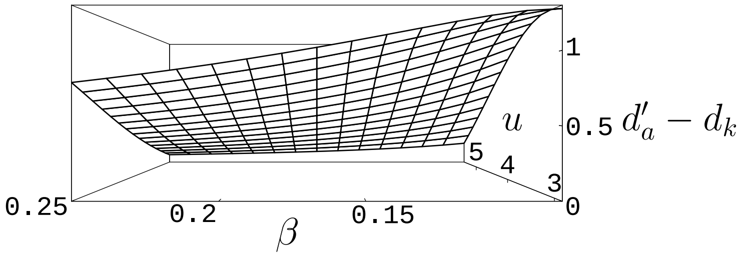

Figure 1: $z = d'_a - d_k,\ n = 64$

4 Concluding Remarks

In this paper, we have shown that particular knapsack cryptosystems using the enumerative source encoding with high density are not secure against the Lagarias–Odlyzko low-density attack. Typical examples are the Chor–Rivest and the Okamoto–Tanaka–Uchiyama cryptosystems, which have been considered to be secure against the density attack. This result can be derived by a more precise analysis of the Lagarias–Odlyzko algorithm than the original one. This implies that 'density' is not a reasonable measure for particular knapsack schemes. Instead, it is interesting to find new measures applicable to these schemes.

[1] While they consider an algebraic number field as a parameter of their system, we restrict our attention to the field of national numbers. This case is optimal with respect to the density, and we can easily modify our analysis to the general case.

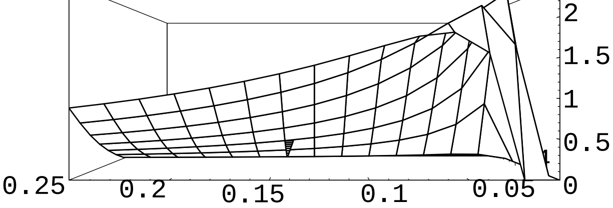

Figure 2: $n = 128$

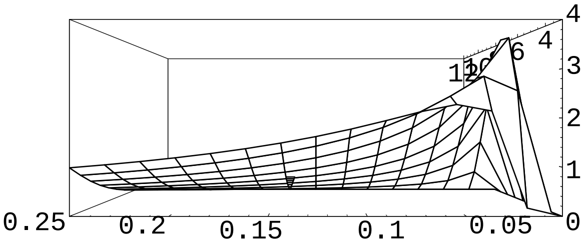

Figure 3: $n = 256$

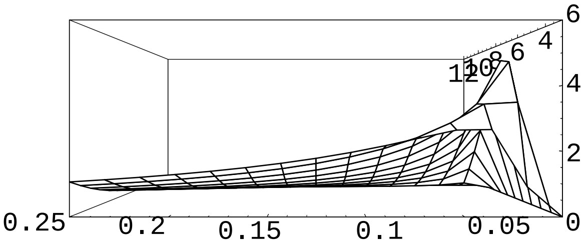

Figure 4: $n = 512$

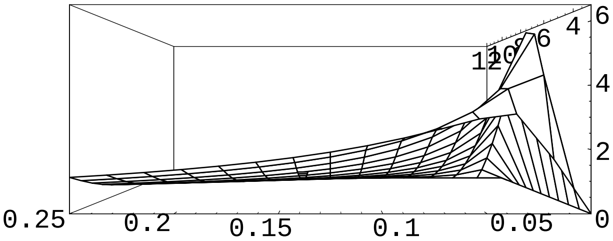

Figure 5: $n = 1024$

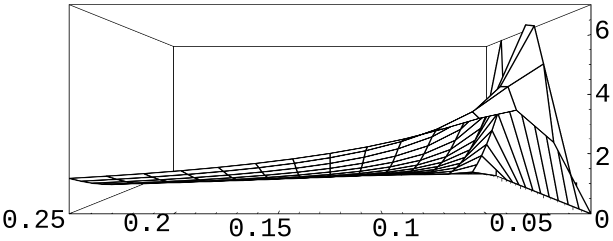

Figure 6: $n = 2048$

References

[1] Brickell, E. F. Breaking iterated knapsacks. In *Advances in Cyrptology: Proceedings of Crypto '84* (1985), Springer, pp. 342–358.

[2] Chor, B., and Rivest, R. L. A knapsack-type public key cryptosystem based on arithmetic in finite fields. In *Advances in Cyrptology: Proceedings of Crypto '84* (1984), vol. 196, Springer, pp. 54–65.

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

[4] Cover, T. M. Enumerative source encoding. In *IEEE Trans. Inform. Theory* (1973), vol. IT-19, pp. 73–77.

[5] Lagarias, J. C., and Odlyzko, A. M. Solving low-density subset sum problems. *Journal of the Association for Computing Machinery 32-1* (1985), 229–246.

[6] Lenstra, A. K., Lenstra Jr., H. W., and Lovász, L. Factoring polynomials with rational coefficients. *Mathematische Annalen 261* (1982), 515–534.

[7] Merkle, R. C., and Hellman, M. E. Hiding information and signatures in trapdoor knapsacks. In *IEEE Trans. Inform. Theory* (1978), vol. IT-24, pp. 525–530.

[8] Okamoto, T., Tanaka, K., and Uchiyama, S. Quantum public-key cryptosystems. In *Advances in Cryptology - CRYPTO 2000* (2000), vol. 1880, Springer, pp. 147–165.

[9] Schnorr, C. P., and Hörner, H. H. Attacking the Chor–Rivest cryptosystem by improved lattice reduction. In *Advances in Cryptology - EUROCRYPT '95* (1995), vol. 921, Springer, pp. 1–12.

[10] Shamir, A. A ploynomial time algorithm for breaking the basic Merkel–Hellman cryptosystem. In *Proceedings of the Twenty-Third Annual Symposium on Foundations of Computer Science* (1982), IEEE, pp. 145–152.

[11] Vaudenay, S. Cryptanalysis of the Chor–Rivest cryptosystem. In *Advances in Cryptology - CRYPTO'98* (1998), vol. 1462, Springer, pp. 243–256.