Adapting Density Attacks to Low-Weight Knapsacks

Phong Q. Nguyễn[1] and Jacques Stern[2]

[1] CNRS & École normale supérieure, DI, 45 rue d'Ulm, 75005 Paris, France.
Phong.Nguyen@di.ens.fr – http://www.di.ens.fr/~pnguyen/

[2] École normale supérieure, DI, 45 rue d'Ulm, 75005 Paris, France.
Jacques.Stern@di.ens.fr – http://www.di.ens.fr/~stern/

Abstract. Cryptosystems based on the knapsack problem were among the first public-key systems to be invented. Their high encryption/decryption rate attracted considerable interest until it was noticed that the underlying knapsacks often had a low density, which made them vulnerable to lattice attacks, both in theory and practice. To prevent low-density attacks, several designers found a subtle way to increase the density beyond the critical density by decreasing the weight of the knapsack, and possibly allowing non-binary coefficients. This approach is actually a bit misleading: we show that low-weight knapsacks do not prevent efficient reductions to lattice problems like the shortest vector problem, they even make reductions more likely. To measure the resistance of low-weight knapsacks, we introduce the novel notion of pseudo-density, and we apply the new notion to the Okamoto-Tanaka-Uchiyama (OTU) cryptosystem from Crypto '00. We do not claim to break OTU and we actually believe that this system may be secure with an appropriate choice of the parameters. However, our research indicates that, in its current form, OTU cannot be supported by an argument based on density. Our results also explain why Schnorr and Hörner were able to solve at Eurocrypt '95 certain high-density knapsacks related to the Chor-Rivest cryptosystem, using lattice reduction.

Keywords: Knapsack, Subset Sum, Lattices, Public-Key Cryptanalysis.

1 Introduction

The knapsack (or subset sum) problem is the following: given a set $\{a_1, a_2, \dots, a_n\}$ of positive integers and a sum $s = \sum_{i=1}^{n} m_i a_i$, where each $m_i \in \{0, 1\}$, recover the m_i's. On the one hand, it is well-known that this problem is NP-hard, and accordingly it is considered to be hard in the worst case. On the other hand, some knapsacks are very easy to solve, such as when the a_i's are the successive powers of two, in which case the problem is to find the binary decomposition of s. This inspired many public-key cryptosystems in the eighties, following the seminal work of Merkle and Hellman [10]:

The Public Key: a set of positive integers $\{a_1, a_2, \dots, a_n\}$.

The Private Key: a method to transform the presumed hard public knapsack into an easy knapsack.

Encryption: a message $m = (m_1, m_2, \ldots, m_n) \in \{0,1\}^n$ is enciphered into $s = \sum_{i=1}^{n} m_i a_i$.

However, with the noticeable exception of the Okamoto-Tanaka-Uchiyama (OTU) quantum knapsack cryptosystem from Crypto '00 [19], all proposed knapsack schemes have been broken (see the survey by Odlyzko [18]), either because of the special structure of the public key (like in [16, 22]) leading to key-recovery attacks, or because of the so-called low-density attacks [6, 3] which allow to decrypt ciphertexts.

The *density* of the knapsack is defined as $d = n/\log_2 A$ where $A = \max_{1\leq i\leq n} a_i$. The density cannot be too high, otherwise encryption would not be injective. Indeed, any subset sum $s = \sum_{i=1}^{n} m_i a_i$ lies in $[0, nA]$, while there are 2^n ways to select the m_i's: if $2^n > nA$, that is, $d > n/(n - \log_2 n)$, there must be a collision $\sum_{i=1}^{n} m_i a_i = \sum_{i=1}^{n} m'_i a_i$, On the other hand, when the density is too low, there is a very efficient reduction from the knapsack problem to the lattice shortest vector problem (SVP): namely, Coster *et al.* [3] showed that if $d < 0.9408\ldots$ (improving the earlier bound $0.6463\ldots$ by Lagarias-Odlyzko [6]), and if the a_i's are chosen uniformly at random over $[0, A]$, then the knapsack problem can be solved with high probability with a single call to a SVP-oracle in dimension n. In practical terms, this means that n must be rather large to avoid lattice attacks (see the survey [17]): despite their NP-hardness, SVP and other lattice problems seem to be experimentally solvable up to moderate dimension. This is why several articles (e.g. [6, 3, 1, 14]) study efficient provable reductions from problems of cryptographic interest to lattice problems such as SVP or the lattice closest vector problem (CVP).

To thwart low-density attacks, several knapsack cryptosytems like Chor-Rivest [2], Qu-Vanstone [16], Okamoto-Tanaka-Uchiyama [19] use in their encryption process a *low-weight* knapsack instead of a random knapsack: $r = \sum_{i=1}^{n} m_i^2$ is much smaller than $n/2$, namely sublinear in n. This means that the message space is no longer $\{0,1\}^n$, but a subset with a special structure, such as the elements of $\{0,1\}^n$ with Hamming weight k, in the case of Chor-Rivest [2] or OTU [19]. Alternatively, it was noticed by Lenstra in [7] that such schemes still work with more general knapsacks where the coefficients are not necessarily 0 or 1: this leads to the *powerline encoding* where the plaintexts are the elements $(m_1, \ldots, m_n) \in \mathbb{N}^n$ such that $\sum_{i=1}^{n} m_i = k$, where again k is much less than $n/2$. With such choices, it becomes possible to decrease the bit-length of the a_i's so as to increase the density d beyond the critical density: a general subset sum $s = \sum_{i=1}^{n} m_i a_i$ may then have several solutions, but one is able to detect the correct one because of its special structure. It was claimed that such knapsack schemes would resist lattice attacks.

Our Results. In this article, we show that low-weight knapsacks are still prone to lattice attacks in theory. Extending earlier work of [6, 3, 20], we provide a general framework to study provable reductions from the knapsack problem to two well-known lattice problems: the shortest vector problem (SVP) and the closest vector problem (CVP). The framework relates in a simple manner the success probability of the reductions to the number of integer points in certain high-dimensional spheres, so that the existence of reductions can be assessed based only on combinatorial arguments, without playing directly with lattices. We notice that this number of integer points can be computed numerically for any realistic choice of knapsacks, which makes it possible to analyze the resistance of any concrete choice of parameters for low-weight knapsack cryptosystems, which we illustrate on the Chor-Rivest cryptosystem. We also provide a simple asymptotic bound on the number of integer points to analyze the theoretical resistance of low-weight knapsack cryptosystems. Mazo and Odlyzko [9] earlier gave sharp bounds in certain cases which are well-suited to usual knapsacks, but not to low-weight knapsacks. As a result, we introduce the so-called *pseudo-density* $\kappa = r \log_2 n / \log_2 A$ (where $r = \sum_{i=1}^{n} m_i^2$) to measure the resistance of low-weight knapsacks to lattice attacks: if κ is sufficiently low, we establish provable reductions to SVP and CVP. This shows that the security of the Okamoto-Tanaka-Uchiyama cryptosystem [19] from Crypto '00 cannot be based on a density argument because its pseudo-density is too low: like NTRU [4], the security requires the hardness of lattice problems. However, we do not claim to break OTU, and we actually believe that this system may be secure with an appropriate choice of the parameters, due to the gap between lattice oracles and existing lattice reduction algorithms, when the lattice dimension is sufficiently high. Our work shows that the density alone is not sufficient to measure the resistance to lattice attacks: one must also take into account the weight of the solution, which is what the pseudo-density does.

Related Work. Omura and Tanaka [20] showed that the Lagarias-Odlyzko reduction [6] could still apply to practical instantiations of the Chor-Rivest and Okamoto-Tanaka-Uchiyama schemes with binary encoding. However, they relied on the counting techniques of Mazo and Odlyzko [9] which are not tailored to low-weight knapsacks. Hence, they could analyze numerically the resistance of any concrete choice of the parameters, but the asymptotical behaviour was not clear. As a result, it was left open to define an analogue of density to low-weight knapsacks, and it was unknown whether or not the reduction could still work when plaintexts were non-binary strings such as in the powerline encoding. Our work shows that more general encodings like the powerline encoding do not rule out lattice attacks either.

Road map. The paper is organized as follows. In Section 2 we provide necessary background on lattices and the number of integer points in high-dimensional spheres. We study reductions from knapsacks to the closest lattice vector problem (CVP) in Section 3, in the case of binary knapsacks and low-weight knapsacks. We then extend those reductions to the shortest lattice vector problem (SVP) in Section 4. We apply our results to the OTU cryptosystem in Section 5, and to the Chor-Rivest cryptosystem in Section 6. Finally, we discuss the significance of our results on the security of low-weight knapsack cryptosystems in Section 7.

Acknowledgements. This work grew out of investigations carried out by the authors under a contract with NTT. We are grateful to NTT for requesting this research and allowing us to publish our results. The preparation of the paper has in part been supported by the Commission of the European Communities through the IST program under contract IST-2002-507932 ECRYPT. We thank Damien Stehlé and the anonymous referees for their helpful comments.

2 Background

2.1 Lattices

Let $\|.\|$ and $\langle ., . \rangle$ be the Euclidean norm and inner product of $\mathbb{R}^n$. We refer to the survey [17] for a bibliography on lattices. In this paper, by the term lattice, we actually mean an integral lattice. An integral lattice is a subgroup of $(\mathbb{Z}^n, +)$, that is, a non-empty subset L of $\mathbb{Z}^n$ which is stable by subtraction: $\mathbf{x} - \mathbf{y} \in L$ whenever $(\mathbf{x}, \mathbf{y}) \in L^2$. The simplest lattice is $\mathbb{Z}^n$. It turns out that in any lattice L, not just $\mathbb{Z}^n$, there must exist linearly independent vectors $\mathbf{b}_1, \ldots, \mathbf{b}_d \in L$ such that:

$$L = \left\{ \sum_{i=1}^{d} n_i \mathbf{b}_i \mid n_i \in \mathbb{Z} \right\}.$$

Any such d-tuple of vectors $\mathbf{b}_1, \ldots, \mathbf{b}_d$ is called a basis of L: a lattice can be represented by a basis, that is, a matrix. Conversely, if one considers d integral vectors $\mathbf{b}_1, \ldots, \mathbf{b}_d \in \mathbb{Z}^n$, the previous set of all integral linear combinations of the $\mathbf{b}_i$'s is a subgroup of $\mathbb{Z}^n$, and therefore a lattice.

The *dimension* of a lattice L is the dimension d of the linear span of L. Since our lattices are subsets of $\mathbb{Z}^n$, they must have a shortest nonzero vector: In any lattice $L \subseteq \mathbb{Z}^n$, there is at least one nonzero vector $\mathbf{v} \in L$ such that no other nonzero lattice vector has a Euclidean norm strictly smaller than that of $\mathbf{v}$. Finding such a vector $\mathbf{v}$ from an arbitrary basis of L is called the *shortest vector problem* (SVP). Another famous lattice problem is the *closest vector problem* (CVP): given a basis of $L \subseteq \mathbb{Z}^n$ and a point $\mathbf{t} \in \mathbb{Q}^n$, find a lattice vector $\mathbf{w} \in L$ minimizing the Euclidean norm of $\mathbf{w} - \mathbf{t}$.

It is well-known that as the dimension increases, CVP is NP-hard and SVP is NP-hard under randomized reductions (see [17, 12] for a list of complexity references). However, in practice, the best lattice reduction algorithms give good results up to moderate dimension: we will discuss this issue in Section 7. This is why it is interesting to study the solvability of various algorithmic problems, when one is given access to a SVP-oracle or a CVP-oracle in moderate dimension. We will call the oracles only once.

2.2 Lattice Points in High-Dimensional Spheres

Following [1, 9], we denote by $N(n, r)$ the number of integer points in the n-dimensional sphere of radius $\sqrt{r}$ centered at the origin: that is, $N(n, r)$ is the number of $(x_1, \ldots, x_n) \in \mathbb{Z}^n$ such that $\sum_{i=1}^{n} x_i^2 \leq r$. Clearly, we have the following induction formula (which was also given in the full version of [1]):

$$N(n,r) = \begin{cases} 1 & \text{if } n = 0 \text{ and } r \geq 0, \\ 0 & \text{if } n = 0 \text{ and } r < 0, \\ \sum_{j=-\lfloor\sqrt{r}\rfloor}^{\lfloor\sqrt{r}\rfloor} N(n-1, r-j^2) & \text{if } n > 0. \end{cases}$$

This allows to compute $N(n,r)$ numerically when n and r are not too large, since the running time is clearly polynomial in (n,r).

When n grows to infinity, sharp estimates of $N(n,r)$ are known when r is proportional to n (see [9]), in which case $N(n,r)$ is exponential in n. Two particular cases are interesting for the knapsack problem: the techniques of Mazo and Odlyzko [9] show that $N(n, n/2) \leq 2^{c_0 n}$ and $N(n, n/4) \leq 2^{c_1 n}$ where $(c_0, c_1) = (1.54724\ldots, 1.0628\ldots)$. Note that $1/c_0 = 0.6463\ldots$ is the critical density of the Lagarias-Odlyzko attack [6], while $1/c_1 = 0.9409\ldots$ is the critical density of the attack of Coster *et al.* [3]. These techniques are very useful when the ratio r/n is fixed and known, but less so for more general choices of n and r.

For low-weight knapsacks, we need to upper bound $N(n,r)$ when r is sublinear in n, in which case the techniques of Mazo and Odlyzko [9] do not seem well-suited. We will use instead the following simple bound:

Lemma 1. *For all $n, r \geq 0$:*

$$N(n,r) \leq 2^r \binom{n+r-1}{r}.$$

Proof. Any vector counted by $N(n,r)$ has at most r non-zero coordinates. Therefore, it suffices to bound the number of integer points with positive coordinates, and to multiply by 2^r to take sign into account. To conclude, the number of integer points with positive coordinates and norm less than $\sqrt{r}$ is clearly bounded by the number K_n^r of combinations of r elements among n with repetition. And it is well-known that $K_n^r = \binom{n+r-1}{r}$. ☐

Corollary 1. *For all $n, r \geq 0$:*

$$N(n,r) \leq \frac{2^r e^{r(r-1)/(2n)} n^r}{r!}.$$

Proof. It suffices to prove that $r!\binom{n+r-1}{r}/n^r \leq e^{r(r-1)/(2n)}$. We have:

$$\begin{aligned} r!\binom{n+r-1}{r}/n^r &= \frac{(n+r-1)(n+r-2)\cdots(n-1)}{n^r} \\ &\leq \prod_{k=1}^{r-1}(1+\frac{k}{n}) \leq \prod_{k=1}^{r-1} e^{k/n} \leq e^{r(r-1)/(2n)} \end{aligned}$$

☐

It follows that if both n and r grow to infinity with a sublinear $r = o(n)$, then $N(n,r) = o(n^r)$ by Stirling's estimate.

3 Reducing Knapsacks to the Closest Vector Problem

In this section, we provide a general framework to reduce the knapsack problem to the closest vector problem. This allows us to easily study the case of low-weight knapsacks, which arguably simplifies the approach of [20] based on [6]. The earlier work [6, 3] only considered reductions to the shortest vector problem, but we start with the closest vector problem because it is simpler to understand, and it gives slightly stronger reductions. We will later adapt those results to the shortest vector problem.

We will distinguish two types of knapsacks. The *binary knapsack* problem is the original knapsack problem: given a set $\{a_1, a_2, \ldots, a_n\}$ of positive integers and a sum $s = \sum_{i=1}^{n} m_i a_i$, where each $m_i \in \{0, 1\}$, recover the m_i's. Because of the powerline encoding, we will also be interested in a more general knapsack problem with non-binary coefficients, which we call the *low-weight knapsack* problem: given a set $\{a_1, a_2, \ldots, a_n\}$ of positive integers and a linear combination $s = \sum_{i=1}^{n} m_i a_i$, where each $m_i \in \mathbb{Z}$ and $r = \sum_{i=1}^{n} m_i^2$ is small, recover the m_i's. The case $r = o(n)$ is of particular interest.

3.1 A General Framework

Solving the knapsack problem amounts to finding a small solution of an inhomogeneous linear equation, which can be viewed as a closest vector problem in a natural way, by considering the corresponding homogeneous linear equation, together with an arbitrary solution of the inhomogeneous equation. Let $s = \sum_{i=1}^{n} m_i a_i$ be a subset sum, where each $m_i \in \{0, 1\}$.

The link between knapsacks and lattices comes from the homogeneous linear equation. Consider indeed the set L of all integer solutions to the homogeneous equation, that is, L is the set of vectors $(z_1, \ldots, z_n) \in \mathbb{Z}^n$ such that:

$$z_1 a_1 + \cdots + z_n a_n = 0. \tag{1}$$

The set L is clearly a subgroup of $\mathbb{Z}^n$ and is therefore a lattice. Its dimension is $n - 1$. It is well-known that a basis of L can be computed in polynomial time from the a_i's (see *e.g.* [16] for one way to do so).

Using an extended gcd algorithm, one can compute in polynomial time integers $y_1, \ldots, y_n$ such that

$$s = \sum_{i=1}^{n} y_i a_i. \tag{2}$$

The y_i's form an arbitrary solution of the inhomogenous equation. Now the vector $\mathbf{v} = (y_1 - m_1, \ldots, y_n - m_n)$ belongs to L. And this lattice vector is fairly close to the vector $\mathbf{t}_1 = (y_1, \ldots, y_n)$ as the coordinates of the difference are the m_i's. The main idea is that by finding the closest vector to $\mathbf{t}_1$ in the lattice L, one may perhaps recover $\mathbf{v}$ and hence the m_i's. The success probability of our reductions will depend in a simple manner on the number of integer points in high-dimensional spheres.

3.2 Binary Knapsacks

In the case of binary knapsacks, the distance between $\mathbf{t}_1$ and $\mathbf{v}$ is roughly $\sqrt{n/2}$. But because $m_i \in \{0,1\}$, the lattice vector $\mathbf{v}$ is even closer to the vector $\mathbf{t}_2 = (y_1 - 1/2, \ldots, y_n - 1/2)$ for which the distance is exactly $\sqrt{n/4}$. It is this simple fact which explains the difference of critical density between the Lagarias-Odlyzko reduction [6] and the reduction by Coster *et al.* [3]. The following results are straightforward:

Lemma 2. *In the case of binary knapsacks, we have:*

1. $\mathbf{v}$ *is a closest vector to* $\mathbf{t}_2$ *in the lattice* L*.*
2. *If* $\mathbf{v}'$ *is a closest vector to* $\mathbf{t}_2$ *in* L*, then* $\|\mathbf{v}' - \mathbf{t}_2\| = \sqrt{n/4}$ *and* $\mathbf{v}'$ *is of the form* $\mathbf{v}' = (y_1 - m'_1, \ldots, y_n - m'_n)$ *where* $s = \sum_{i=1}^{n} m'_i a_i$ *and* $m'_i \in \{0,1\}$*.*

Proof. The key observation is that elements of the lattice have integer coordinates and that each coordinate contributes to the distance to $\mathbf{t}_2$ by at least $1/2$. □

This gives a deterministic polynomial-time reduction from the binary knapsack problem to the closest vector problem (CVP) in a lattice of dimension $n-1$: this reduction was sketched in the survey [17], and can be viewed as a variant of an earlier reduction by Micciancio [11], who used a different lattice whose dimension was n, instead of $n-1$ here.

Thus, a single call to a CVP-oracle in an $(n-1)$-dimensional lattice automatically gives us a solution to the binary knapsack problem, independently of the value of the knapsack density, but this solution may not be the one we are looking for, unless the unicity of the solution is guaranteed. One particular case for which the unicity is guaranteed is Merkle-Hellman: more generally, for any *traditional* knapsack cryptosystem such that the set of plaintexts is the whole $\{0,1\}^n$ without decryption failures, a single call to a CVP-oracle is sufficient to decrypt.

It is nevertheless interesting to know when one can guarantee the unicity of the solution for general knapsacks. But if for instance some a_i is a subset sum of other a_j's where $j \in J$, then clearly, all knapsacks involving only a_i and a_ℓ's where $\ell \notin J$ may also be decomposed differently using the a_j's where $j \in J$. This means that to guarantee unicity of solutions in a general knapsack, we may only hope for probabilistic statements, by considering random knapsacks where the a_i's are assumed to be chosen uniformly at random in $[0, A]$:

Theorem 1. *Let* $(m_1, \ldots, m_n) \in \{0,1\}^n$*. Let* $a_1, \ldots, a_n$ *be chosen uniformly and independently at random in* $[0, A]$*. Let* $s = \sum_{i=1}^{n} m_i a_i$*. Let* L *and the* y_i*'s be defined by (1) and (2). Let* $\mathbf{c}$ *be a vector in* L *closest to the vector* $\mathbf{t}_2 = (y_1 - 1/2, \ldots, y_n - 1/2)$*. Then the probability that* $\mathbf{c}$ *is not equal to* $(y_1 - m_1, \ldots, y_n - m_n)$ *is less than* $(2^n - 1)/A$*.*

Proof. By Lemma 2, $\mathbf{c}$ is of the form $\mathbf{c} = (y_1 - m'_1, \ldots, y_n - m'_n)$ where $s = \sum_{i=1}^{n} m'_i a_i$ and $m'_i \in \{0,1\}$. If $\mathbf{c}$ is not equal to $(y_1 - m_1, \ldots, y_n - m_n)$, then

$\mathbf{m}' = (m'_1, \ldots, m'_n) \neq \mathbf{m} = (m_1, \ldots, m_n)$. But:

$$\sum_{i=1}^{n} (m_i - m'_i) a_i = 0. \tag{3}$$

Since $\mathbf{m} \neq \mathbf{m}'$, there exists i_0 such that $m_{i_0} \neq m'_{i_0}$. For any choice of $(a_i)_{i \neq i_0}$, there exists a unique choice of a_{i_0} satisfying (3), since $m_{i_0} - m'_{i_0} = \pm 1$. It follows that for a given $\mathbf{m}' \neq \mathbf{m}$, the probability that $(y_1 - m'_1, \ldots, y_n - m'_n)$ is equal to $\mathbf{c}$ is less than $1/A$. We conclude since the number of $\mathbf{m}'$ is $2^n - 1$. □

This shows that when the density $d = n/\log_2 A$ is < 1, there is with high probability a unique solution, and this solution can be obtained by a single call to a CVP-oracle in dimension $n - 1$.

3.3 Low-Weight Knapsacks

We showed that the hidden vector $\mathbf{v} \in L$ related to the knapsack solution was relatively close to two target vectors $\mathbf{t}_1$ and $\mathbf{t}_2$. In fact, $\mathbf{v}$ was a lattice vector closest to $\mathbf{t}_2$: the distance was $\sqrt{n/4}$. In the general binary case, this was better than $\mathbf{t}_1$ for which the distance was expected to be $\sqrt{n/2}$, provided that the Hamming weight of the knapsack was roughly $n/2$. But if the Hamming weight k is much smaller than $n/2$, then the distance between $\mathbf{m}$ and $\mathbf{t}_1$ is only $\sqrt{k}$, which is much less than $\sqrt{n/4}$. We obtain the following general result regarding low-weight knapsacks (not necessarily binary):

Theorem 2. *Let $\mathbf{m} = (m_1, \ldots, m_n) \in \mathbb{Z}^n$. Let $a_1, \ldots, a_n$ be chosen uniformly and independently at random in $[0, A]$. Let $s = \sum_{i=1}^{n} m_i a_i$. Let L and the y_i's be defined by (1) and(2). Let $\mathbf{c}$ be a vector in L closest to the vector $\mathbf{t}_1 = (y_1, \ldots, y_n)$. Then the probability that $\mathbf{c}$ is not equal to $(y_1 - m_1, \ldots, y_n - m_n)$ is less than $N(n, \|\mathbf{m}\|^2)/A$.*

Proof. By definition, $\mathbf{c}$ is of the form $\mathbf{c} = (y_1 - m'_1, \ldots, y_n - m'_n)$ where $s = \sum_{i=1}^{n} m'_i a_i$ and $m'_i \in \mathbb{Z}$. Let $\mathbf{m}' = (m'_1, \ldots, m'_n)$. Because $\mathbf{c}$ cannot be farther from $\mathbf{t}_1$ than $\mathbf{v}$, $\|\mathbf{m}'\| \leq \|\mathbf{m}\|$. If $\mathbf{c}$ is not equal to $(y_1 - m_1, \ldots, y_n - m_n)$, then $\mathbf{m}' \neq \mathbf{m} = (m_1, \ldots, m_n)$: there exists i_0 such that $m_{i_0} \neq m'_{i_0}$. For any choice of $(a_i)_{i \neq i_0}$, there exists at most one choice of a_{i_0} satisfying (3). It follows that for a given $\mathbf{m}' \neq \mathbf{m}$, the probability that $(y_1 - m'_1, \ldots, y_n - m'_n)$ is the closest vector is less than $1/A$. We conclude since the number of $\mathbf{m}'$ is less than $N(n, \|\mathbf{m}\|^2)$, as $\|\mathbf{m}'\| \leq \|\mathbf{m}\|$. □

Note that $N(n, \|\mathbf{m}\|^2)$ can be evaluated numerically from Section 2.2, so that one can bound the failure probability for any given choice of the parameters.

We saw that $\mathbf{t}_1$ was better than $\mathbf{t}_2$ with low-weight knapsacks, but the choice $\mathbf{t}_1$ can be improved if $k = \sum_{i=1}^{n} m_i \neq 0$, which is the case of usual knapsacks where all the m_i's are positive. Consider indeed $\mathbf{t}_3 = (y_1 - k/n, y_2 - k/n, \ldots, y_n - k/n)$. Then $\|\mathbf{v} - \mathbf{t}_3\|^2 = \|\mathbf{m}\|^2 - k^2/n$ which is less than $\|\mathbf{v} - \mathbf{t}_1\|^2 = \|\mathbf{m}\|^2$. By replacing $\mathbf{t}_1$ with $\mathbf{t}_3$ in Theorem 2, the result becomes:

Theorem 3. *Let* $\mathbf{m} = (m_1, \ldots, m_n) \in \mathbb{Z}^n$ *and* $k = \sum_{i=1}^n m_i$. *Let* $a_1, \ldots, a_n$ *be chosen uniformly and independently at random in* $[0, A]$. *Let* $s = \sum_{i=1}^n m_i a_i$. *Let* L *and the* y_i*'s be defined by (1) and(2). Let* $\mathbf{c}$ *be a vector in* L *closest to the vector* $\mathbf{t}_3 = (y_1 - k/n, \ldots, y_n - k/n)$. *Then the probability that* $\mathbf{c}$ *is not equal to* $(y_1 - m_1, \ldots, y_n - m_n)$ *is less than* $N(n, \|\mathbf{m}\|^2 - k^2/n)/A$.

If $k = \sum_{i=1}^n m_i$ is proportional to n, Theorem 3 yields a significant improvement over Theorem 2: for instance, if we consider a binary random knapsack for which $k \approx n/2$, Theorem 3 involves $N(n, n/4)$ instead of $N(n, n/2)$ for Theorem 2, which is exactly the difference between the critical densities of the Lagarias-Odlyzko reduction [6] and the reduction by Coster *et al.* [3]. However, in the case of low-weight knapsacks where $k = o(n)$, the improvement becomes marginal, as k^2/n is then negligible with respect to $\|\mathbf{m}\|^2$. To simplify the presentation and the discussion, we will therefore rather consider Theorem 2.

4 Reducing Knapsacks to the Shortest Vector Problem

In the previous section, we established reductions from knapsack problems (binary and low-weight) to the closest vector problem. The original lattice attacks [6, 3] on knapsacks only considered reductions to the shortest vector problem (SVP), not to CVP. In this section, we show that our reductions to CVP can be adapted to SVP, thanks to the well-known embedding or (homogenization) method introduced by Kannan (see [5, 12, 13]), which tries to transform an $(n-1)$-dimensional CVP to an n-dimensional SVP. In general, the embedding method is only heuristic, but it can be proved in the special case of knapsack lattices. This is interesting from a practical point of view, because CVP is often solved that way.

We adapt Theorem 2 to SVP. Again, we let $s = \sum_{i=1}^n m_i a_i$. Let L be the lattice defined by (1), and let the $y_i's$ be defined by (2). Let $(\mathbf{b}_1, \ldots, \mathbf{b}_{n-1})$ be a basis of L. We embed L into the n-dimensional lattice L' spanned by $(1, y_1, \ldots, y_n) \in \mathbb{Z}^{n+1}$ and the $n-1$ vectors of the form $(0, \mathbf{b}_i) \in \mathbb{Z}^{n+1}$. We let $\mathbf{m}' = (1, m_1, \ldots, m_n) \in \mathbb{Z}^{n+1}$. By definition, $\mathbf{m}' \in L'$ and its norm is relatively short. The following result lowers the probability that $\mathbf{m}'$ is the shortest vector of L'.

Theorem 4. *Let* $\mathbf{m} = (m_1, \ldots, m_n) \in \mathbb{Z}^n$. *Let* $a_1, \ldots, a_n$ *be chosen uniformly and independently at random in* $[0, A]$. *Let* $s = \sum_{i=1}^n m_i a_i$. *Let* L', $\mathbf{m}'$ *and the* y_i*'s be defined as previously. Let* $\mathbf{s}$ *be a shortest non-zero vector in* L'. *Then the probability that* $\mathbf{s}$ *is not equal to* $\pm\mathbf{m}'$ *is less than*

$$(1 + 2(1 + \|\mathbf{m}\|^2)^{1/2}) N(n, \|\mathbf{m}\|^2)/A.$$

Proof. By definition of L', $\mathbf{s}$ is of the form $\mathbf{s} = (r, ry_1 - z_1, \ldots, ry_n - z_n)$ where $r \in \mathbb{Z}$, and $(z_1, \ldots, z_n) \in L$. Since $\mathbf{s}$ is a shortest vector:

$$\|\mathbf{s}\|^2 \leq \|\mathbf{m}'\|^2 = 1 + \|\mathbf{m}\|^2. \tag{4}$$

It follows that $r^2 \leq 1 + \|\mathbf{m}\|^2$. Let $u_i = ry_i - z_i$ and $\mathbf{u} = (u_1, \ldots, u_n)$. We have $\|\mathbf{u}\| \leq \|\mathbf{s}\|$. Notice that:

$$\sum_{i=1}^{n} (u_i - rm_i)a_i = 0. \tag{5}$$

We distinguish two cases. If $r = 0$, then $\mathbf{u} \neq 0$, and it follows that the probability of (5) being satisfied for a given $\mathbf{u} \neq 0$ is less than $1/A$. And the number of possible $\mathbf{u}$ is bounded by $N(n, \|\mathbf{m}\|^2)$. Otherwise, $r \neq 0$, and there are at most $2(1 + \|\mathbf{m}\|^2)^{1/2}$ possible values for r. If $\mathbf{s} \neq \pm\mathbf{m}'$, we claim that there exists i_0 such that $u_{i_0} - rm_{i_0} \neq 0$, in which case the probability that (5) is satisfied is less than $1/A$. Otherwise, $\mathbf{u} = r\mathbf{m}$: if $|r| > 1$, this would imply that $\|\mathbf{u}\| \geq \|\mathbf{m}\|$, and $\mathbf{s}$ would not be shorter than $\mathbf{m}'$; else $r = \pm 1$, and $\mathbf{u} = \pm\mathbf{m}$ which contradicts $\mathbf{s} \neq \pm\mathbf{m}'$. This concludes the proof. □

Theorem 4 provides essentially the same bound on the success probability as Theorem 2, because $\|\mathbf{m}\|$ is negligible with respect to $N(n, \|\mathbf{m}\|^2)$. This means that in the case of low-weight knapsacks, there is no significant difference between the CVP and SVP cases.

Theorem 4 can be viewed as a generalization of the Lagarias-Odlyzko result [6]. Indeed, if we consider a binary knapsack of Hamming weight $\leq n/2$ (which we may assume without loss of generality), then the failure probability is less than

$$(1 + 2(1 + n/2)^{1/2})N(n, n/2)/A.$$

Since $N(n, n/2) \leq 2^{c_0 n}$ where $c_0 = 1.54724\ldots$ (see Section 2), it follows that the failure probability of the reduction to SVP is negligible provided that the density $d = n/\log_2 A$ is strictly less than $1/c_0 = 0.6463\ldots$, which matches the Lagarias-Odlyzko result [6].

We omit the details but naturally, the improvement of Theorem 3 over Theorem 2 can be adapted to Theorem 4 as well: $N(n, \|m\|^2)$ would decrease to $N(n, \|m\|^2 - k^2/n)$ where $k = \sum_{i=1}^{n} m_i$, provided that one subtracts k/n to both y_i and m_i in the definition of L' and $\mathbf{m}'$. In the particular case of binary knapsacks, this matches the result of Coster *et al.* [3]: because $N(n, n/4) \leq 2^{c_1 n}$ where $c_1 = 1.0628\ldots$, the failure probability would be negligible provided that the knapsack density is less than $1/c_1 = 0.9409\ldots$ Whereas there was almost no difference between the CVP reduction and the SVP reduction for low-weight knapsacks, there is a difference in the case for binary knapsacks: in Theorem 1, the critical density was 1 and not $1/c_1$. And that would not have changed if we had transformed the CVP-reduction of Theorem 1 (instead of that of Theorem 3) into a probabilistic reduction to SVP. This is because Lemma 2 used in Theorem 1 (but not in Theorem 3) has no analogue in the SVP setting, which explains why the result with a CVP-oracle is a bit stronger than with a SVP-oracle: there are more parasites with SVP.

In other words, the framework given in Section 3 revisits the SVP reductions of Lagarias-Odlyzko [6] and Coster et al. [3]. By applying the embedding technique, we obtain the same critical densities when transforming our CVP reductions of Theorem 2 and 3 into SVP reductions.

5 Application to the OTU Cryptosystem

In this section, we apply the results of Sections 2, 3 and 4 to the Okamoto-Tanaka-Uchiyama cryptosystem [19] from Crypto 2000.

5.1 Description of OTU

The OTU cryptosystem is a knapsack cryptosystem where the knapsack has a hidden structure based on discrete logarithms like the Chor-Rivest scheme [2], but where no information on the DL group leaks, thwarting attacks like [22]. The key generation of OTU requires the extraction of discrete logarithms: if quantum computers are available, one can apply Shor's quantum algorithm, otherwise one uses groups with a special structure (e.g. groups of smooth order) so that DL is tractable.

The knapsack $(a_1, \dots, a_n)$ used by OTU has a special structure. Let $A = \max_{1 \leq i \leq n} a_i$. To allow decryption, it turns out that A is such that $A \geq p^k$ for some integers $p, k > 1$, and p is such that there are at least n coprime numbers $\leq p$, which implies that $p \geq n$, and therefore $A \geq n^k$, and $\log_2 A$ is at least linear in k. The OTU scheme allows two kinds of encoding:

- The binary encoding, where the plaintexts are all $(m_1, \dots, m_n) \in \{0,1\}^n$ such that $\sum_{i=1}^n m_i = k$.
- The powerline encoding [7], where the plaintexts are all $(m_1, \dots, m_n) \in \mathbb{N}^n$ such that $\sum_{i=1}^n m_i = k$.

There is no concrete choice of parameters proposed in [19]. However, it was pointed out on page 156 of [19] that the choice $k = 2^{(\log n)^c}$ where c is a constant < 1 would have interesting properties. We will pay special attention to that case since it is the only asymptotical choice of k given in [19], but we note from the discussion in [19, Section 3.4] that the scheme could tolerate larger values of k, up to maybe a constant times $n/\log n$. Perhaps the main drawback with larger values of k is the keysize, as the storage of the knapsack is $\Omega(nk)$ bits, which is then essentially quadratic if $k = n/\log n$. What is clear is that k is at most $O(n/\log n)$: indeed the density in OTU is $O(n/(k \log n))$, and the density must be lower bounded by a constant > 0 to ensure the hardness of the knapsack, which implies that $k = O(n/\log n)$. This means that we should study two cases: the suggested case $k = 2^{(\log n)^c}$ where c is a constant < 1, and the extreme case $k = O(n/\log n)$.

5.2 Resistance to Low-Density Attacks

The parameter A can be chosen as small as $O(p^k)$ and p can be as small as $n \log n$. For the suggested case $k = 2^{(\log n)^c}$, we have $\log A = O(k \log p) = o(n)$. It follows that the usual density $d = n/\log_2 A$ grows to infinity, which is why it was claimed in [19] that OTU prevents usual lattice attacks [6, 3]. However, this density argument is misleading because the weight k is sublinear in n.

Let $\mathbf{m} = (m_1, \ldots, m_n)$ and $s = \sum_{i=1}^n m_i a_i$. Theorems 4 and 2 provide efficient reductions from knapsacks to SVP and CVP, provided that $N(n, \|\mathbf{m}\|^2)$ is negligible with respect to A.

With the binary encoding, we have $\|\mathbf{m}\|^2 = k$, and therefore $N(n, \|\mathbf{m}\|^2) = N(n, k)$. We know that due to the choice of k in OTU (even in the extreme case), we have $k = o(n)$ with k growing to infinity. Corollary 1 then implies that $N(n, k) = o(n^k)$, and therefore $N(n, k)/A = o(1)$ since $A \geq n^k$. Hence Theorems 4 and 2 provide efficient reductions (with success probability asymptotically close to 1) to SVP and CVP in dimension n, provided that $k = o(n)$, which is a necessary requirement for OTU.

We now show that the powerline encoding does not significantly improve the situation, even though a plaintext $\mathbf{m}$ with the powerline encoding only satisfies $k \leq \|\mathbf{m}\|^2 \leq k^2$. If $\|\mathbf{m}\|^2$ was close to k^2, rather than k, Corollary 1 on $N(n, \|\mathbf{m}\|^2)$ would not allow us to conclude, because n^{k^2} would dominate A. The following result shows that $\|\mathbf{m}\|^2$ is on the average much closer to k, as in the binary encoding:

Theorem 5. *There exists a computable constant $\alpha > 0$ such that the following holds. Let $1 \leq k \leq n$ and $y = (k-1)/n$. Let $\mathbf{m} = (m_1, \ldots, m_n) \in \mathbb{N}^n$ be chosen uniformly at random such that $\sum_{i=1}^n m_i = k$. Then the expected value of $\|\mathbf{m}\|^2$ satisfies:*

$$E(\|\mathbf{m}\|^2) \leq k(1 + \alpha y).$$

Proof. As in the proof of Lemma 1, let K_n^k denote the number of combinations of k elements among n with repetition: $K_n^k = \binom{n+k-1}{k} = \binom{n+k-1}{n-1}$. We have:

$$\begin{aligned} E(\|\mathbf{m}\|^2) = nE(m_i^2) &= n \sum_{x=1}^{k} x^2 \frac{K_{n-1}^{k-x}}{K_n^k} \\ &= n \sum_{x=1}^{k} x^2 \frac{k(k-1)\cdots(k-x+1) \times (n-1)}{(n+k-1)(n+k-2)\cdots(n+k-x-1)}. \end{aligned}$$

Let:

$$s(n, x, k) = n(n-1)x^2 \frac{k(k-1)\cdots(k-x+1)}{(n+k-1)(n+k-2)\cdots(n+k-x-1)},$$

so that $E(\|\mathbf{m}\|^2) = \sum_{x=1}^k s(n, x, k)$. We will see that the first term dominates in this sum:

$$s(n, 1, k) = \frac{n(n-1)k}{(n+k-1)(n+k-2)} \leq k.$$

We now bound $s(n, x, k)$ for all $2 \leq x \leq k$:

$$\begin{aligned} s(n, x, k) &\leq kx^2 \frac{(k-1)(k-2)\cdots(k-x+1)}{(n+k-1)(n+k-2)\cdots(n+k-x+1)} \\ &= kx^2 \prod_{u=k-x+1}^{k-1} \frac{u}{n+u} \leq kx^2 \left(\frac{k-1}{n+k-1}\right)^{x-1} \end{aligned}$$

$$\leq kx^2 \left(\frac{y}{1+y}\right)^{x-1} \text{ with } y = \frac{k-1}{n}.$$

Hence, by separating the first two terms in the sum:

$$E(\|\mathbf{m}\|^2) \leq k \left(1 + \frac{4y}{1+y} + \sum_{x=3}^{k} x^2 \left(\frac{y}{1+y}\right)^{x-1}\right).$$

Because $1 \leq k \leq n$, we have $0 \leq y < 1$ and $0 \leq y/(1+y) < 1/2$. Thus, we only need to bound the series:

$$f(y) = \sum_{x=3}^{\infty} x^2 \left(\frac{y}{1+y}\right)^{x-1}.$$

A short derivative computation shows that for any $0 \leq z < 1/2$, the function $x \mapsto x^2 z^{x-1}$ decreases over $x \geq 3$, because $2 + 3\ln(1/2) < 0$. Therefore, letting $z = y/(1+y)$, we obtain for all $k > 1$:

$$f(y) \leq \int_2^{\infty} x^2 z^{x-1} dx = \left[\frac{z^{x-1}}{\ln z}\left(x^2 - \frac{2x}{\ln z} + \frac{2}{\ln^2 z}\right)\right]_2^{\infty} = \frac{-z}{\ln z}\left(4 - \frac{4}{\ln z} + \frac{2}{\ln^2 z}\right).$$

Since $z \leq 1/2$, it follows that one can compute an absolute constant $\beta > 0$ such that for all $k > 1$, $f(y) \leq \beta z$, which in fact also holds when $k = 1$, that is, $z = 0$. Hence for all $1 \leq k \leq n$:

$$E(\|\mathbf{m}\|^2) \leq k \left(1 + \frac{4y}{1+y} + \beta z\right) \leq k(1 + (4+\beta)y).$$

This concludes the proof with $\alpha = 4 + \beta$. □

When $k = o(n)$, we have $y = o(1)$ and the upper bound becomes $k(1 + \alpha y) = k(1 + o(1))$, which already shows that with the powerline encoding, the expected value of $\|\mathbf{m}\|^2$ is essentially k, rather than k^2. This suggests that $N(n, \|\mathbf{m}\|^2)$ will on the average still be negligible with respect to A. But Theorem 5 allows us to give a sharper estimate. In the extreme case of OTU, we have $k = O(n/\log n)$ growing to infinity, so $y = O(1/\log n)$ and the upper bound becomes $r = k(1 + O(1/\log n))$. By Corollary 1:

$$N(n, r)/A \leq \frac{2^r e^{r(r-1)/(2n)} n^r}{r! n^k}.$$

Here, $r^2/n = kO(n/\log n)(1 + O(1/\log n))/n = O(k/\log n)$ therefore:

$$2^r e^{r(r-1)/(2n)} = O(1)^k.$$

And $n^r = n^{k(1+O(1/\log n))} = n^k \times (n^{O(1/\log n)})^k \leq n^k \times O(1)^k$. Hence:

$$N(n, r)/A \leq \frac{O(1)^k}{r!} = o(1).$$

Thus, the reductions of Theorems 4 and 2 succeed with overwhelming probability even with the powerline encoding, even if the extreme choice of k in OTU is considered. This question was left open in [20].

Although we believe that the OTU cryptosystem may be secure with an appropriate choice of the parameters, our results indicate that in its current form, it cannot be supported by an argument based on density that would protect the system against a single call to an SVP oracle or a CVP oracle.

5.3 The Pseudo-Density

We now explain why in the case of low-weight knapsacks, Theorems 4 and 2 suggest to replace the usual density $d = n/\log_2 A$ by a pseudo-density defined by $\kappa = r\log_2 n/\log_2 A$, where r is an upper bound on $\|\mathbf{m}\|^2$, $\mathbf{m}$ being the knapsack solution.

Theorems 4 and 2 showed that a low-weight knapsack could be solved with high probability by a single call to a SVP-oracle or a CVP-oracle, provided that $N(n,r)/A$ was small. Corollary 1 shows that:

$$N(n,r)/A \leq \frac{2^r e^{r(r-1)/(2n)}}{r!} \times \frac{n^r}{A}.$$

The left-hand term $2^r e^{r(r-1)/(2n)}/r!$ tends to 0 as r grows to ∞, provided that $r = O(n)$. The right-hand term n^r/A is $2^{r\log_2 n - \log_2 A}$. This shows that if the pseudo-density κ is ≤ 1, then the right-hand term will be bounded, and therefore the low-weight knapsack can be solved with high probability by a single call to either a SVP-oracle or a CVP-oracle. On the other hand, if the pseudo-density κ is larger than 1, it will not necessarily mean that the previous upper bound does not tend to zero, as there might be some compensation between the left-hand term and the right-hand term.

Consider for instance the case of OTU with binary encoding. For any choice of k, the pseudo-density $\kappa = k\log_2 n/\log_2 A$ is ≤ 1 because $A \geq n^k$ due to decryption requirements. Therefore there is a reduction to SVP and CVP with probability asymptotically close to 1. On the other hand, if we consider the powerline encoding with an extreme case of k, the pseudo-density becomes $\kappa = k(1+O(1/\log n))\log_2 n/\log_2 A \leq 1+O(1/\log n)$ which could perhaps be slightly larger than 1. Nevertheless, the computation of the previous section showed that $N(n,r)/A$ was still $o(1)$. Thus, the pseudo-density is a good indicator, but it may not suffice to decide in critical cases.

6 Application to the Chor-Rivest Cryptosystem

The Chor-Rivest cryptosystem [2] is another low-weight knapsack cryptosystem, which survived for a long time until Vaudenay [22] broke it, for all the parameter choices proposed by the authors in [2]. Vaudenay used algebraic techniques specific to the Chor-Rivest scheme, which do not apply to OTU. His attack recovers the private key from the public key. Schnorr and Hörner [21] earlier tried to decrypt Chor-Rivest ciphertexts by solving the underlying low-weight knapsack using an improved lattice reduction method which they introduced. They succeeded for certain choices of moderate parameters, but failed for the parameter choices proposed in [2]. Despite the fact that the Chor-Rivest scheme is broken, it is an interesting case with respect to lattice attacks, and this is why we apply our results to this scheme.

6.1 Description

We give a brief description of the Chor-Rivest cryptosystem [2]. One selects a small prime q and an integer k such that one can compute discrete logarithms in $\mathrm{GF}(q^k)$. One computes the discrete logarithms $b_1, \dots, b_q \in \mathbb{Z}_{q^k-1}$ of certain well-chosen elements in $\mathrm{GF}(q^k)$, to ensure decryption. The elements of the knapsack are $a_i = b_i + d$ where d is an integer chosen uniformly at random in $\mathbb{Z}_{q^k-1}$. The set of plaintexts is the subset of all $(m_1, \dots, m_q) \in \{0,1\}^q$ having Hamming weight k, and the encryption of $(m_1, \dots, m_q)$ is:

$$s = \sum_{i=1}^{q} a_i m_i \pmod{q^k - 1}.$$

The public key consists of the q, k and the a_i's.

Strictly speaking, Chor-Rivest involves a modular knapsack problem (modulo $q^k - 1$), rather than the initial knapsack problem. The density of the Chor-Rivest knapsack is $d = q/(k \log q)$, which can therefore be rather high for appropriate choices of q and k. But all our results on the knapsack problem we have discussed can be adapted to the modular knapsack problem. First of all, notice that a modular knapsack can be transformed into a basic knapsack if one can guess the hidden multiple of $q^k - 1$ involved, that is, if one knows the integer ℓ such that:

$$s + \ell(q^k - 1) = \left(\sum_{i=1}^{q} a_i m_i \right).$$

Clearly, ℓ can be exhaustively searched, and it is very close to k. In the worst-case for our reductions to lattice problems, the number of oracle calls will increase very slightly.

Alternatively, one can adapt the lattice used in our framework. Consider a modular knapsack $s = \sum_{i=1}^{n} a_i m_i \pmod{A}$. We replace the lattice L defined by (1) by the set L of vectors $(z_1, \dots, z_n) \in \mathbb{Z}^n$ such that:

$$z_1 a_1 + \cdots + z_n a_n \equiv 0 \pmod{A}. \tag{6}$$

The set L is a subgroup of $\mathbb{Z}^n$ and is therefore a lattice. Its dimension is n, rather than $n - 1$. It is again well-known that a basis of L can be computed in polynomial time. This time, we compute in polynomial time integers $y_1, \dots, y_n$ such that

$$s \equiv \sum_{i=1}^{n} y_i a_i \pmod{A}. \tag{7}$$

All of our results, such as Theorems 1–4, can then be adapted to modular knapsacks provided some obvious minor changes, which we omit. For instance, in the statements of Theorems 1–4, the uniform distribution must be over $[0, A[$, and we let $s = \sum_{i=1}^{n} a_i m_i \pmod{A}$. Naturally, equations (1) and (2) must be replaced respectively by equations (6) and (7).

6.2 Application

By definition, the pseudo-density of the Chor-Rivest knapsack (with binary encoding) is $\kappa = k \log_2 q / \log_2(q^k) = 1$. We thus conclude that the low-weight knapsack problems arising from the Chor-Rivest cryptosystem can be efficiently reduced to SVP and CVP with probability close to 1. In retrospect, it is therefore not surprising that Schnorr and Hörner [21] were able to solve certain Chor-Rivest knapsacks using lattice reduction.

Concretely, we can even compute upper bounds on the failure probability of the reduction for the parameters proposed in [2] and the ones used in [21], using numerical values of $N(n, r)$, as explained in Section 2.2. The numerical results are summarized in Tables 1 and 2. Thus, if one had access to SVP-oracles or CVP-oracles in dimension roughly 200–250, one could decrypt Chor-Rivest ciphertexts with overwhelming probability for its proposed parameters.

Table 1. Application to the Chor-Rivest parameters proposed in [2].

Value of (q, k)	(197,24)	(211,24)	(256,25)	(243,24)
Value of $N(q, k)/q^k$	2^{-57}	2^{-57}	2^{-60}	2^{-57}

Table 2. Application to the Chor-Rivest parameters attacked in [21].

Value of (q, k)	(103,12)	(151,16)
Value of $N(q, k)/q^k$	2^{-18}	2^{-29}

7 Impact on the Security of Low-Weight Knapsack Cryptosystems

We have established efficient provable reductions from the low-weight knapsack problem to two well-known lattice problems: SVP and CVP. However, we do not claim to break low-weight knapsack cryptosystems like OTU. This is because there is an experimental and theoretical gap between lattice oracles for SVP/CVP and existing lattice reduction algorithms (see [17] for a list of references), as the lattice dimension increases. The state-of-the-art in lattice reduction suggests that exact SVP and CVP can only be solved up to moderate dimension, unless the lattice has exceptional properties (such as having one extremely short non-zero vector compared to all the other vectors).

To roughly estimate the hardness of SVP/CVP in a m-dimensional lattice of volume V, lattice practitioners usually compare $V^{1/m}\sqrt{m}$ with a natural quantity related to the expected solution: for SVP, the quantity is the norm of the expected shortest vector, while for CVP, it is the distance between the target vector and the lattice. If the ratio is not large, it means that the solution is not exceptionally small: SVP and CVP become intractable in practice if the dimension is sufficiently high. In the case of a knapsack defined by integers $a_1, \ldots, a_n$, the work of [16] on the so-called orthogonal lattices show as a simple particular case that the lattice L defined by (1) has volume $V = (\sum_{i=1}^{n} a_i^2)^{1/2} / \gcd(a_1, \ldots, a_n)$. Thus, with overwhelming probability, $V \approx A = \max_i a_i$. Since the dimension of L is $n-1$, we need to consider $V^{1/(n-1)} \approx 2^{(\log_2 A)/(n-1)} \approx 2^{1/d}$ where d is the usual knapsack density. The quantity is thus $V^{1/(n-1)}\sqrt{n-1} \approx 2^{1/d}\sqrt{n}$. When dealing with a low-weight knapsack of weight $r = \sum_{i=1}^{n} m_i^2$, this quantity is not particularly large compared to the quantity $\sqrt{r}$ corresponding ot the solution, unless r is extremely small. This indicates that by taking a sufficiently high dimension n and a not too small r (which is also important to avoid simple dimension reduction methods like [8]), the corresponding lattice problems should be hard.

One may wonder how to select the lattice dimension to guarantee the hardness of SVP and CVP in practice. Current experimental records in lattice computations seem to depend on the type of lattices. For instance, Schnorr and Hörner [21], using what is still the best lattice reduction algorithm known in practice, failed to decrypt Chor-Rivest ciphertexts for its suggested parameters, which correspond to a lattice dimension around 200–250. Bleichenbacher and Nguyen [1] reported similar problems with a dense 160-dimensional lattice. On the other hand, Nguyen [13] broke the GGH-challenge in dimension 350, but not in dimension 400. The record computation for breaking the NTRU cryptosystem [4] is a SVP computation in dimension 214 by May (see [8]), while the smallest NTRU parameter currently proposed corresponds to a 502-dimensional lattice. Thus, in order to propose concrete parameters for OTU, it would be useful to gather experimental data with the best reduction algorithms known (keeping track of recent development such as [15]). Besides, SVP and CVP instances arising from knapsack problems could serve as a useful benchmark to test and design new lattice reduction algorithms.

References

1. D. Bleichenbacher and P. Q. Nguyễn. Noisy polynomial interpolation and noisy Chinese remaindering. In *Proc. of Eurocrypt '00*, volume 1807 of *LNCS*. IACR, Springer-Verlag, 2000.
2. B. Chor and R.L. Rivest. A knapsack-type public key cryptosystem based on arithmetic in finite fields. *IEEE Trans. Inform. Theory*, 34, 1988.

3. M.J. Coster, A. Joux, B.A. LaMacchia, A.M. Odlyzko, C.-P. Schnorr, and J. Stern. Improved low-density subset sum algorithms. *Comput. Complexity*, 2:111–128, 1992.
4. J. Hoffstein, J. Pipher, and J.H. Silverman. NTRU: a ring based public key cryptosystem. In *Proc. of ANTS III*, volume 1423 of *LNCS*, pages 267–288. Springer-Verlag, 1998. Additional information and updates at `http://www.ntru.com`.
5. R. Kannan. Minkowski's convex body theorem and integer programming. *Math. Oper. Res.*, 12(3):415–440, 1987.
6. J. C. Lagarias and A. M. Odlyzko. Solving low-density subset sum problems. *Journal of the Association for Computing Machinery*, January 1985.
7. H.W. Lenstra, Jr. On the Chor-Rivest knapsack cryptosystem. *J. of Cryptology*, 3:149–155, 1991.
8. A. May and J. Silverman. Dimension Reduction Methods for Convolution Modular Lattices. In *Cryptography and Lattices – Proc. of CALC*, volume 2146 of *LNCS*. Springer-Verlag, 2001.
9. J. E. Mazo and A. M. Odlyzko. Lattice points in high-dimensional spheres. *Monatsh. Math.*, 110:47–61, 1990.
10. R. Merkle and M. Hellman. Hiding information and signatures in trapdoor knapsacks. *IEEE Trans. Inform. Theory*, IT-24:525–530, September 1978.
11. D. Micciancio. The hardness of the closest vector problem with preprocessing. *IEEE Trans. Inform. Theory*, 47(3):1212–1215, 2001.
12. D. Micciancio and S. Goldwasser. *Complexity of lattice problems: A cryptographic perspective*. Kluwer Academic Publishers, Boston, 2002.
13. P. Q. Nguyễn. Cryptanalysis of the Goldreich-Goldwasser-Halevi cryptosystem from Crypto '97. In *Proc. of Crypto '99*, volume 1666 of *LNCS*, pages 288–304. IACR, Springer-Verlag, 1999.
14. P. Q. Nguyễn and I. E. Shparlinski, *The Insecurity of the Digital Signature Algorithm with Partially Known Nonces*, Journal of Cryptology, vol. 15, no. 3, pp. 151–176, Springer, 2002.
15. P. Q. Nguyễn and D. Stehlé. Floating-Point LLL Revisited. In *Proc. of Eurocrypt '05*, volume 3494 of *LNCS*. IACR, Springer-Verlag, 2005.
16. P. Q. Nguyễn and J. Stern. Merkle-Hellman revisited: a cryptanalysis of the Qu-Vanstone cryptosystem based on group factorizations. In *Proc. of Crypto '97*, volume 1294 of *LNCS*, pages 198–212. IACR, Springer-Verlag, 1997.
17. P. Q. Nguyễn and J. Stern. The two faces of lattices in cryptology. In *Cryptography and Lattices – Proc. of CALC*, volume 2146 of *LNCS*. Springer-Verlag, 2001.
18. A. M. Odlyzko. The rise and fall of knapsack cryptosystems. In *Cryptology and Computational Number Theory*, volume 42 of *Proc. of Symposia in Applied Mathematics*, pages 75–88. A.M.S., 1990.
19. T. Okamoto, K. Tanaka, and S. Uchiyama. Quantum Public-Key Cryptosystems. In *Proc. of Crypto '00*, *LNCS*. Springer-Verlag, 2000.
20. K. Omura and K. Tanaka. Density Attack to the Knapsack Cryptosystems with Enumerative Source Encoding. In *IEICE Trans. Fundamentals*, vol. E84-A, No. 1, January (2001).
21. C. P. Schnorr and H. H. Hörner. Attacking the Chor-Rivest cryptosystem by improved lattice reduction. In *Proc. of Eurocrypt '95*, volume 921 of *LNCS*, pages 1–12. IACR, Springer-Verlag, 1995.
22. S. Vaudenay Cryptanalysis of the Chor-Rivest Cryptosystem. In *Journal of Cryptology*, vol. 14 (2001), pp 87-100.