Lattice analysis on MiNTRU problem.

Changmin Lee[1] and Alexandre Wallet[2]

[1] Univ. Lyon, EnsL, UCBL, CNRS, Inria, LIP, F-69342 Lyon Cedex 07, France
changmin.lee@ens-lyon.fr
[2] NTT Corporation, Japan
alexandre.wallet.th@ntt.co.jp

Abstract. In ASIACRYPT 2019, Genise *et al.* describe [GGH+19] a new somewhat homomorphic encryption scheme. The security relies on an inhomogeneous and non-structured variant of the NTRU assumption that they call MiNTRU. To allow for meaningful homomorphic computations, overstretched parameters are used, but an analysis of their new assumption against the state-of-the-art attack of Kirchner and Fouque [KF17] for overstretched modulus is not provided. We show that the parameters of [GGH+19] do not satisfy the desired security by actually conducting the known analysis. We also report a successful break of the MiNTRU assumption for smallest set of parameters in around 15 hours of computations while they are claimed to reach 100 bits of security.

1 Introduction

Introduced by Hoffstein, Pipher and Silverman [HPS98], the NTRU problem is informally the following: given a polynomial $h := f/g \bmod q \in \mathbb{Z}_q[X]/\Phi$ where $f, g \in \mathbb{Z}[X]$ are secret with small coefficients and Φ is an integer polynomial, recover the pair (f, g). For specific polynomials Φ such as (but not limited to) powers-of-two cyclotomic polynomials, it is believed to be (quantumly) hard, and its variants have been popular choices to design efficient post-quantum schemes: among others, NIST round 2 candidates (KEMs [ZCH+19, BCLv19], signatures [DDLL13, PFH+19]), an IBE scheme [DLP14], multilinear maps [GGH13, LSS14, ACLL15] and homomorphic encryption schemes [LATV12, BLLN13].

The so-called "overstretched" variant uses a huge modulus q compared to the dimension $2n$ of the underlying lattices. An application of this variant was used to construct homomorphic encryption [LATV12, BLLN13] and a candidate of multilinear map [GGH13, LSS14, ACLL15]. However, this largeness of q induced a disastrous security loss for such schemes. Cheon, Jeong and Lee [CJL16] and Albrecht, Bai and Ducas [ABD16] independently presented *the subfield attack* on the overstretched NTRU problem, which was already a huge blow to the security level for the proposed parameters. These attacks made strong use of the algebraic structure of the underlying number field. Soon after, Kirchner and Fouque [KF17] showed that, in fact, the attack boils down to pure lattice reduction of a well-chosen sublattice of the NTRU lattice $\Lambda_q := \{(u, v) \in \mathbb{Z}^{2n} : vh - u = 0 \bmod q\}$. The crux of the attack is to observe that (f, g) is a very short vector of this lattice, and that together with its Galois conjugates, it spans a rank n sublattice $\mathcal{L}$ of unusually small volume. Because the volume $\mathrm{Vol}(\Lambda_q) = q^n$ is very large in the overstretched case, this gap between volumes has to be compensated in some way: a strong enough lattice reduction over

$$\mathbf{B}_{NTRU} = \begin{pmatrix} q\mathbf{I}_n & \mathbf{0} \\ h & \mathbf{I}_n \end{pmatrix}$$

will find vectors shorter than the minima of the "orthogonal complement" of $\mathcal{L}$. In other words, these vectors will necessarily belong to the lattice spanned by the secret (f, g). Moreover, because the Gram-Schmidt orthogonalization preserves volume during basis reduction, such short vectors are likely to be detected in large enough sublattices. This intuition can be made more formal using a lemma of Pataki and Tural [PT08] guaranteeing that the product of the smallest Gram-Schmidt is smaller than the volume of any sublattice of $\mathcal{L}(\mathbf{B}_{NTRU})$, combined with the Geometric Series Assumption.[3] This allowed Kirchner and Fouque to improve the practical efficiency of the attack, and to actually break some overstretched NTRU schemes. Additionally, it showed that the algebraic structure provided by the cyclotomic ring had little impact on the concrete security in the overstretched case. The only benefit of the structure is to actually ensure that, as long as one short vector is obtained, then an entire full-rank sublattice is deduced by action of the Galois conjugates; on an anecdotal level, it also helps in estimating the volume of the small sublattice. But ultimately, only geometric properties and the existence of a "large rank but very small volume" sublattice are core to the attack.

In ASIACRYPT 2019, Genise *et al.* describe [GGH+19] a new somewhat homomorphic encryption scheme. To prove the semantic security of their design, the authors relies on an inhomogeneous and non-structured variant of the NTRU assumption that they call MiNTRU. Let $\mathbf{G} = [\mathbf{I}_n | \dots | 2^{\log q - 1}\mathbf{I}_n] \in \mathbb{Z}_q^{n \times m}$ be the so-called gadget matrix, with $m = n \log q$. Then, the variant of the MiNTRU problem considered by [GGH+19] is the following: given $\mathbf{A} := \mathbf{S}^{-1} \cdot (\mathbf{G} - \mathbf{E}) \bmod q \in \mathbb{Z}_q^{n \times m}$ where $\mathbf{S} \in \mathbb{Z}_q^{n \times n}$ and $\mathbf{E} \in \mathbb{Z}_q^{n \times m}$ are random binary matrices, recover the pair $(\mathbf{S}, \mathbf{E})$.[4] For their most efficient set of parameters, the authors claim 100 bits of security. However, in order to support meaningful homomorphic computations, this scheme also uses overstretched modulus. Yet, the authors do not provide any analysis of the impact of the sublattice attack against their scheme.

Our contribution. We show that the current choice of practical parameters of MiNTRU problem do not give semantic security by actually recovering $\mathbf{S}$ from $\mathbf{A} = \mathbf{S}^{-1}(\mathbf{G} - \mathbf{E}) \bmod q$. As expected, it amounts to applying several sublattice attacks for suitable parameters, the rest of the attack having negligible cost overall. For the smallest parameter sets proposed by [GGH+19], we ran the attack successfully in around 15 hours of computations with fplll/BKZ 2.0 [NS09, MSV09, CN11] in Sagemath on a single core of a personal laptop. As it involves different lattices, the full lattice phase can be parallelized easily and would recover the encryption key $\mathbf{S}$ in essentially this amount of time. In practice, an attacker may not have access to an exact MiNTRU sample as the encryption algorithm draws an additional small noise, so $\mathbf{E}$ is no more a random binary matrix. Still, this already strongly suggests that the overstretched parameters really hinders the security guarantee of the scheme. Future directions would be to mount the attack against the encryption scheme itself. It seems promising, as the encryption noise is still a low norm matrix: there is still a small volume but high rank sublattice in every NTRU-lattice related to a ciphertext.

[3] This heuristic states that after lattice reduction, the Gram-Schmidt norms of the outputted basis decrease geometrically.

[4] To be precise, they rely on a decisional version of this problem.

We expect that a stronger lattice reduction would be necessary, but considering that our attack only needs a block-size of 20 and a target lattice of dimension essentially $n/4$, they should stay in the realm of practical computations. Overall, we hope to make it clear that overstretched parameters should be avoided when designing NTRU-based schemes.

Organization After some preliminaries, we give for the sake of completeness a quick reminder of the analysis for the lattice phase of our attack. After describing the second phase of the attack, we provide experimental results for smaller parameters in the overstretched ranges. We conclude with some more details on the cost of the attack for the parameters of [GGH+19], and discuss potential directions for attacking the encryption scheme itself.

2 Preliminaries

We recall standard heuristic assumption about lattices, and state some useful technical lemmas.

2.1 Lattices

A lattice $\mathcal{L}$ is a discrete subgroup of $\mathbb{R}^m$. It is usually represented by a basis, that is, a set of linearly independent vectors $\mathbf{b}_1, \ldots, \mathbf{b}_k$. The integer k is called the rank of $\mathcal{L}$. The Gaussian Heuristic gives an estimate of the shortest vector of a given lattice. For "random enough" lattices and as the dimension grows, it tends to be acceptably accurate, although it is not difficult to find counter-examples. Indeed, a large class of the NTRU lattices used in cryptography are such lattices where the shortest vector is way below the Gaussian Heuristic. This NTRU lattices of this work are no exception, although we will apply the heuristic to projections of these lattices.

Heuristic 1 (Gaussian heuristic). For any lattice $\mathcal{L}$ of rank k, we have

$$\lambda_1(\mathcal{L}) = \sqrt{\frac{k}{2\pi \mathrm{e}}} \cdot \mathrm{Vol}(\mathcal{L})^{1/k}.$$

The Geometric Series Assumption is nowadays a standard heuristic assumption to predict the behaviour of lattice block-reduction algorithms. It has been backed-up by extensive experimental results [Ajt06], and expresses that the norms of the Gram-Schmidt vectors after reduction decrease in a geometric manner.

Heuristic 2 (Geometric Series Assumption (GSA)). Let $\mathcal{L}$ be a rank k lattice with basis $\mathbf{b}_1, \ldots, \mathbf{b}_k$. After execution of BKZ with block-size β, the norms of the Gram-Schmidt vectors satisfy

$$\|\mathbf{b}_i^*\| = \delta_\beta^2 \cdot \|\mathbf{b}_{i+1}^*\|,$$

for $1 \leq i \leq k-1$, and where $\delta_\beta = \left(\frac{\beta}{2\pi \mathrm{e}}(\pi\beta)^{1/\beta}\right)^{1/(2(\beta-1))}$.

The quantity δ_β is known as the β-root Hermite factor. It is used to get estimations on Gram-Schmidt norms.

Lemma 2.1 (Heuristic). *Let $k \geq 1$ be an integer, and $\mathbf{B} \in \mathbb{Z}^{2k\times 2k}$ be a basis. Let $\mathbf{b}_1^*, \ldots, \mathbf{b}_{2k}^*$ be the rows of the Gram-Schmidt orthogonalization of $\mathbf{B}$ after performing lattice reduction in block-size β. If the Geometric Series Assumption holds, we have*

$$\delta_\beta^{-k(3k-1)} \cdot \|\mathbf{b}_1\|^k = \prod_{i=1}^{k} \|\mathbf{b}_{k+i}^*\|,$$

where δ_β is the β-root Hermite Factor.

Proof. This is successive applications of the GSA to $\prod_{i=1}^{k} \|\mathbf{b}_{k+i}^*\|$. □

Lemma 2.2 (Pataki-Tural). *Let $\mathcal{L}$ be a full rank lattice in $\mathbb{R}^n$ and $\mathbf{b}_1, \ldots, \mathbf{b}_n$ be a basis of $\mathcal{L}$. For any rank $d \leq n$ sublattice $\mathcal{L}'$ of $\mathcal{L}$, we have*

$$\min_{\substack{S\subset[n]\\|S|=d}} \prod_{i\in S} \|\mathbf{b}_i^*\| \leq \operatorname{Vol} \mathcal{L}'.$$

2.2 Binary matrices

We collect two useful results on random binary matrices.

Lemma 2.3. *Let $m \geq n \geq 1$ be integers and $\mathbf{X} \hookleftarrow \{0,1\}^{n\times m}$ be a random binary matrix with i.i.d. entries. When $\mathbf{X}$ has full rank, then we have $\mathbb{E}[\operatorname{Vol}(\mathcal{L}(\mathbf{X}))] \leq (\frac{m}{2})^n$.*

Proof. For a Bernoulli vector $\mathbf{x}$ of length m with i.i.d entries, we have $\mathbb{E}[\|\mathbf{x}\|^2] = m/2$. If $\mathbf{X}$ has full rank, $\mathbf{X}\mathbf{X}^t$ is positive definite. By Hadamard's inequality and independence of the $\mathbf{x}_i$'s, we have $\mathbb{E}[\det \mathbf{X}\mathbf{X}^t] \leq (\frac{m}{2})^n$. The result follows from Jensen's inequality or positivity of the variance. □

For example, if $m = 2n$, $\operatorname{Vol}(\mathcal{L}(\mathbf{X}))$ should be no more than $n^{n/2}$. However this upper bound seemed experimentally quite loose. When needed, we will use the more accurate (but experimental) upper bound of $(n/2)^{n/2}$. We also use an asymptotic estimate on the smallest singular value of a random binary square matrix (which are in particular subgaussians). More precise results are known but they give more than we actually need for our attack. Indeed, the next statement is verified pretty well in experiments, which is enough for us.

Proposition 2.4 (Adapted from [Ver07]). *Let $\mathbf{S}$ be an $n \times n$ Bernoulli matrix, and let s be its smallest singular values. Then with high probability, we have $s \approx 1/\sqrt{n}$. In particular, when the latter event happens, we have $\|\mathbf{S}^{-1}\|_\infty \approx n$.*

2.3 The MiNTRU problem

We recall the Genise *et al.*'s formulation of the MiNTRU problem. For more details, we refer to the original article [GGH+19]. First choose integers $n \leq m, q$ and some centered distribution χ which should outputs elements way smaller than q with overwhelming probability. Let $\mathbf{G} = [\mathbf{I}_n | \ldots | 2^{\log q - 1} \cdot \mathbf{I}_n] \in \mathbb{Z}_q^{n \times m}$ be the gadget matrix of dimensions $n \times m$. The MiNTRU distribution samples $\mathbf{S} \hookleftarrow \chi^{n \times n}$ until it is invertible modulo q, $\mathbf{E} \hookleftarrow \chi^{n \times m}$, and then outputs a sample

$$\mathbf{C} = \mathbf{S}^{-1}(\mathbf{G} - \mathbf{E}) \bmod q.$$

The computational variant is to recover $(\mathbf{S}, \mathbf{E})$ from $\mathbf{C}$, and the decisional variant is to distinguish $\mathbf{C}$ from a random matrix in $\mathbb{Z}_q^{n \times m}$. We note that the semantic security of their homomorphic encryption scheme is implied by hardness of the decisional-MiNTRU assumption, and that the decisional variant reduces to the computational variant. In particular, we would like to underline that ciphertexts have a bit more noise than a plain MiNTRU sample; we do not give more details as we are not interested in ciphertexts in this work, rather the MiNTRU problem in itself.

Practical parameters: the concrete choice for χ in [GGH+19] is the uniform binary distribution, so we only consider this choice further in this article. The other relevant parameters are the pair $(n, q) \in \{(1024, 2^{42}), (4096, 2^{111}), (32768, 2^{883})\}$, and we only care about the smallest pair $(1024, 2^{42})$. Still, we note that all these choices have $q = 2^{\alpha\sqrt{n}}$ for some $\alpha > 1$, which qualifies as an overstretched regime of parameters.

3 Lattice based analysis

3.1 Overview

In this section, we describe an attack algorithm that recovers $\mathbf{S}$ from a given $\mathbf{C} = \mathbf{S}^{-1}(\mathbf{G} - \mathbf{E}) \bmod q$ obtained from the MiNTRU distribution. This immediately allows to distinguish the MiNTRU distribution from random, and thus makes the security proof void for the concrete parameters. The attack runs in two phases. First, lattice reduction is performed over (possibly several) sublattices of the NTRU-like lattice $\Lambda_q(\mathbf{C}_0) = \{(\mathbf{u}, \mathbf{v}) \in \mathbb{Z}^{2n} : \mathbf{u}\mathbf{C}_0 - \mathbf{v}\}$, where $\mathbf{C}_0$ is the first $n \times n$ block of the cipher. This is the most costly part, and we will analyze heuristically its behaviour in the next section. If we let $\mathbf{E}_0$ be the first $n \times n$ block of $\mathbf{E}$, observe that $\mathcal{L}([\mathbf{S}, \mathbf{I}_n - \mathbf{E}_0])$ is a rank n sublattice of $\Lambda_q(\mathbf{C}_0)$. At the end of this lattice phase, we recover a matrix $\mathbf{X}' = [\mathbf{S}', \mathbf{E}'] \in \mathbb{Z}^{n \times 2n}$ with short rows, generating a full-rank sublattice of $\mathcal{L}([\mathbf{S}, \mathbf{I}_n - \mathbf{E}])$. In particular, we know that there is $\mathbf{T} \in \mathbb{Z}^{n \times n}$ such that $\mathbf{S}' = \mathbf{T} \cdot \mathbf{S}$. If $\mathbf{S}$ is a binary matrix, we expect by Lemma 2.4 that the rows of $\mathbf{T}$ have a size not too much larger than those of $\mathbf{S}'$. Therefore, if we have $\mathbf{S}' \cdot \mathbf{C}_i \bmod q = \mathbf{T} \cdot (2^i \mathbf{I}_n - \mathbf{E}_i)$ for each i's, then we recover $\mathbf{T} = \lceil 2^{-i} \mathbf{S}' \cdot \mathbf{C}_i \rfloor$ for some i, by rounding each entries to the closest integer.

3.2 Analysis of the lattice phase

We keep the notation of the previous section, and focus on the lattice $\Lambda_q(\mathbf{C}_0)$. When $\mathbf{C}_0$ is invertible (which happens with high probability), it can be checked that it admits the basis matrix

$$\mathbf{B} := \begin{pmatrix} q \cdot \mathbf{I}_n & \mathbf{0}_n \\ \mathbf{C}_0 & \mathbf{I}_n \end{pmatrix},$$

and, as mentioned before, each row of $\mathbf{X} := [\mathbf{S}, \mathbf{I}_n - \mathbf{E}_0]$ belongs to this lattice by construction.

When the dimension of $\mathbf{B}$ is large, it is unlikely that a lattice reduction algorithm on the full matrix $\mathbf{B}$ will terminate fast enough to qualify as an efficient attack. Thus we follow the Kirchner and Fouque approach [KF17] in order to reduce the dimension of the problem to a practical range. The main idea is to extract a suitable submatrix and to perform a lattice reduction algorithm on the submatrix. More precisely, the basis matrix $\mathbf{B}$ can be divided into blocks as

$$\mathbf{B} = \left(\begin{array}{c|cc|c} q\cdot\mathbf{I}_{n-k} & 0 & 0 & 0 \\ \hline 0 & q\cdot\mathbf{I}_k & 0 & 0 \\ \mathbf{C}_{00} & \mathbf{C}_{01} & \mathbf{I}_k & 0 \\ \hline \mathbf{C}_{10} & \mathbf{C}_{11} & 0 & \mathbf{I}_{n-k} \end{array}\right),$$

where $\mathbf{C}_{ij}$ is the corresponding block matrix of the matrix $\mathbf{C}_0$, and we consider the central lower triangular submatrix

$$\mathbf{B}' = \begin{pmatrix} q\cdot\mathbf{I}_k & \mathbf{0}_k \\ \mathbf{C}_{01} & \mathbf{I}_k \end{pmatrix}.$$

We let $\mathbf{b}'_1, \ldots, \mathbf{b}'_k$ be the basis obtained by performing lattice reduction in block size β over $\mathbf{B}'$.

We heuristically assume that the output basis follow the Geometric Series Assumption (GSA). It implies in particular that the k last Gram-Schmidt vectors are the smallest ones. By Pataki-Tural lemma, this product is bounded by the volume of any rank k sublattice $\mathcal{L}$. Combining with Lemma 2.1, we have for such a lattice that

$$\delta_\beta^{-k(3k-1)} \cdot \|\mathbf{b}'_1\|^k \leq \operatorname{Vol}\mathcal{L}.$$

We now argue that a lattice $\mathcal{L}(\mathbf{B}')$ includes a k-rank sublattice $\mathcal{L}$ such that $\operatorname{Vol}(\mathcal{L}) \leq \operatorname{Vol}(\mathcal{L}(\mathbf{X}))$, and uses it as an upper bound of $\|\mathbf{b}'_1\|$ in the equation above. The Hermite normal form of the matrix $\mathbf{X} \in \mathbb{Z}^{n\times 2n}$ is likely to be

$$\left(\begin{array}{cc|ccc} \mathbf{x}_{11} & \mathbf{x}_{12} & \det(\mathbf{I}_n - \mathbf{E}_0) & 0 & 0 \\ \mathbf{x}_{21} & \mathbf{x}_{22} & \mathbf{x}_{23} & \mathbf{I}_{k-1} & 0 \\ \mathbf{x}_{31} & \mathbf{x}_{32} & \mathbf{x}_{33} & 0 & \mathbf{I}_{n-k} \end{array}\right),$$

where the $\mathbf{x}_{ij}$'s are the corresponding block matrices. In particular, each block matrix $\mathbf{x}_{i2}$ has k columns. Considering the $k \times 2k$ submatrix

$$\mathbf{X}' := \left(\begin{array}{c|cc} \mathbf{x}_{12} & \det(\mathbf{I}_n - \mathbf{E}_0) & 0 \\ \mathbf{x}_{22} & \mathbf{x}_{23} & \mathbf{I}_{k-1} \end{array}\right),$$

we then see that 1) $\operatorname{Vol}(\mathcal{L}(\mathbf{X}')) \leq \operatorname{Vol}(\mathcal{L}(\mathbf{X}))$ and 2) each row of $\mathbf{X}'$ are included in the lattice $\mathcal{L}(\mathbf{B}')$. To sum-up, we must have $\|\mathbf{b}'_1\| \leq \delta_\beta^{3k-1}\operatorname{Vol}(\mathcal{L}(\mathbf{X}'))^{1/k}$.

On the other hand, let $\mathcal{L}^\perp$ be the orthogonal projection of $\mathcal{L}(\mathbf{B}')$ into the space orthogonal to the one spanned by $\mathbf{X}'$. The Gaussian Heuristic in $\mathcal{L}^\perp$ gives us

$$\lambda_1(\mathcal{L}^\perp)^k = \left(\frac{k}{2\pi e}\right)^{k/2} \cdot \frac{q^k}{\operatorname{Vol}(\mathcal{L}(\mathbf{X}'))}.$$

If $\|\mathbf{b}'_1\| < \lambda_1(\mathcal{L}^\perp)$, then it means that $\mathbf{b}'_1 \in \mathcal{L}(\mathbf{X}')$. To understand when this happens, we assume by contradiction that $\|\mathbf{b}'_1\| \geq \lambda_1(\mathcal{L}^\perp)$. Combining everything so far, this implies that

$$\delta_\beta^{-k(3k-1)} \cdot \left(\frac{k}{2\pi \mathrm{e}}\right)^{k/2} \cdot q^k \leq \mathrm{Vol}(\mathcal{L}(\mathbf{X}'))^2, \tag{3.1}$$

and we are now looking for (k, β) violating this condition. For such a pair, we can conclude that the last k entries in $\mathbf{b}'_1$ (appropriately padded with zeros) give a vector in $\mathcal{L}(\mathbf{S})$. Observe that the smaller $\mathrm{Vol}(\mathcal{L}(\mathbf{X}'))$ is compared to q, the smaller k and β will be.

In practice: In practice we start by selecting the $2k$ central rows of $\mathbf{B}$ and perform lattice reduction. Next, we repeat this process by selecting the $n-k, \ldots, n$-th rows and the $n+k+1, \ldots, n+2k$-th rows instead, and so on until the full matrix has been covered. This gives several linearly independent lattice vectors in $\mathcal{L}(\mathbf{S})$. If we do not have enough to span a full rank sublattice, we can continue with the next $n \times n$ block of $\mathbf{C}$, or taking other subsets of rows, or just re-run lattice reduction on the initial choice. It is clear that all these lattice steps can be parallelized, so it boils down to see the practical cost of the first lattice reduction. As claimed, we end this phase with a matrix $\mathbf{S}'$ generating a full-rank sublattice of $\mathcal{L}(\mathbf{S})$. Experimentally, the behaviour of lattice reduction is in fact even better: reducing the central subalttice for k, β large enough, one finds in fact k short vectors among the k first vectors of the reduced basis, the k next ones being far greater (as the overall volume should be preserved).

We now explain how to compute parameters (k, β) which satisfy the condition 3.1. To check the condition, we need to estimate $\mathrm{Vol}(\mathcal{L}(\mathbf{X}'))$. In the worst case, $\mathrm{Vol}(\mathcal{L}(\mathbf{X}'))$ is the same as $\mathrm{Vol}(\mathcal{L}(\mathbf{X}))$, so we replace it with $\mathrm{Vol}(\mathcal{L}(\mathbf{X}))$ and use the experimental estimate of $(n/2)^{n/2}$ (see the discussion below Lemma 2.3). In other words, we aim at satisfying the following condition:

$$\left(\frac{k}{2\pi \mathrm{e}}\right)^{k/2} \cdot q^k \geq \delta_\beta^{k(3k-1)} \cdot (n/2)^n. \tag{3.2}$$

There are known estimate [GN08, Che13] for the root Hermite factor of the LLL and BKZ algorithm with block size β. For given parameters of n and q, we first select β and then search for the smallest k that satisfies the condition 3.2.

Asymptotically, taking logarithms with $k = n\log(n/2)/(3\log\delta_\beta)$ and simplifying, it can be rewritten as the mildly weaker

$$\log q \geq 3k \cdot \log\delta_\beta + n/k \cdot \log(n/2) = \sqrt{12n\log\delta_\beta \cdot \log(n/2)}.$$

It implies that when $\log q = \alpha \cdot \sqrt{n}$ with $\alpha > 1$, $\log\delta_\beta = \mathcal{O}(\frac{1}{\log(n/2)})$ is enough to find such a short vector.

3.3 Recovering the secret key

At this stage, we assume that $\mathbf{S}' = \mathbf{T} \cdot \mathbf{S}$ is known. Our next goal is to recover $\mathbf{T}$, from which $\mathbf{S}$ is easily deduced. Noting that the size of $2^i\mathbf{T}$ may be larger than q, we claim that we can compute $\mathbf{T} \cdot (2^i \mathbf{I}_n - \mathbf{E}_i)$ over the integers for all i. First, we note that $\mathbf{D}_0 := \mathbf{S}'\mathbf{C}_0 \bmod q = \mathbf{T} \cdot (\mathbf{I}_n - \mathbf{E}_0)$ holds over the integers. Observe that the matrix $2\mathbf{E}_i - \mathbf{E}_{i+1}$ is small for all i, so that we can also compute $\mathbf{D}_i := \mathbf{S}'(2\mathbf{C}_i - \mathbf{C}_{i+1}) \bmod q = \mathbf{T} \cdot (2\mathbf{E}_i - \mathbf{E}_{i+1})$. We then readily check that $\mathbf{T} \cdot (2^i \mathbf{I}_n - \mathbf{E}_i) = \sum_{0 \le j \le i} 2^{i-j} \mathbf{D}_j$ over the integers too, giving our claim. Lastly, recall that $r = \log q - 1$ is an integer. According to result of the Section 3 and with Lemma 2.4, we know that $\|\mathbf{S}'\|_\infty \le \frac{n}{\sqrt{2\pi e}} \cdot \frac{q}{\mathrm{Vol}(\mathcal{L}(\mathbf{X}))^{1/k}}$, so that we expect that $\|\mathbf{T}\mathbf{E}_r\|_\infty \le \|\mathbf{S}'\|_\infty \|\mathbf{S}^{-1}\|_\infty \|\mathbf{E}_r\|_\infty \le q/4$. Therefore, rounding the entries of $\mathbf{T} \cdot (2^r \mathbf{I}_n - \mathbf{E}_r)/2^r$ recovers $\mathbf{T}$.

4 Experiments and practical attack

In Table 1, we give experimental results for several smaller parameter sets. When the block-size is 2, the LLL algorithm was used instead of the BKZ algorithm. The parameter $2k$ represents the number of rows of the matrices used in the lattice reduction phase. In all experiments we succeeded in recovering the secret key $\mathbf{S}$.

According to our computations, the BKZ algorithm with a block size of 20 allows a successful attack in dimension $n = 2^8$ and $n = 2^9$. However, in our experiments, the LLL algorithm was in fact enough to recover $\mathbf{S}$. One can see that the LLL algorithm overperforms.

$\log n$	$\log q$	block size β	# of rows, $2k$	$\max \log(\|\mathbf{U} \cdot \mathbf{S}\|_\infty)$	$\max \log(\|\mathbf{U}\|_\infty)$
6	22	2	24	7.2479	6.9773
7	27	2	50	9.2192	10.4888
8	32	20(2)	100	12.4571	11.7507
9	37	20(2)	216	15.2833	13.4098

Table 1. Experimental results of the several parameters of MiNTRU problem.

The practical attack: The smallest parameters of [GGH+19] are $n = 2^{10}$, $q = 2^{42}$. As already mentioned, this means that the MiNTRU distribution selects χ to be the uniform binary distribution, so that $\mathbf{S}, \mathbf{E}$ are random binary matrices of dimensions $n \times n$ and $n \times n \log q$, respectively. We note that a security level of ≈ 100 bits is claimed for the encryption scheme for these parameters. We started by computing a given sample of the MiNTRU distribution, and run lattice reduction using the cipher and taking $k = 280$ and $\beta = 20$. We used Sagemath 9.0 and its version of BKZ 2.0 included in fplll (version 0.5.1). This is a floating-point implementation, and we selected a precision of 180 bits since else, the Gram-Schmidt computations tended to go in “infinite loop in Babai” state. After around 15 hours of computations on a personal laptop, we obtained k somewhat short vectors with a log-norm of roughly 22, all in the lattice $\mathcal{L}(\mathbf{S})$. The code for this attack can be found at `http://github.com/awallet/Overstretched`. Observe that with 4 cores, more than n such vectors can be found by this approach. The log-norm of the transformation matrix $\mathbf{T}$ is then expected to be way below $q/4$, so we are essentially assured that the full attack will work out. This means that these parameters do not give a pseudorandom MiNTRU distribution. The other sets of parameters are not as practical, and the gap between

q and n is even worse, so that the attack is likely to succeed for smaller k's and β's (relatively to the overall dimension). Overall, we conclude that the security proof of the scheme is void for the concrete parameters that are proposed.

Possible future directions It could be interesting to attack directly the encryption and possibly target the larger parameters. The strategy is exactly the same, but the larger noise will impact negatively the choice of (k, β) for the attack to work. We believe that it is still be possible to recover the secret encryption key with moderate computing resources and time, because the added noise is still very small. This strongly suggests that one should avoid overstretched modulus when designing primitive relying on an NTRU-like assumption, structured or not.

Acknowledgments We want to thank Damien Stehlé for suggesting to look at this scheme as well as for fruitful discussions.

References

ABD16. Martin Albrecht, Shi Bai, and Léo Ducas. A subfield lattice attack on overstretched ntru assumptions. In *Annual International Cryptology Conference*, pages 153–178. Springer, 2016.

ACLL15. Martin R Albrecht, Catalin Cocis, Fabien Laguillaumie, and Adeline Langlois. Implementing candidate graded encoding schemes from ideal lattices. In *International Conference on the Theory and Application of Cryptology and Information Security*, pages 752–775. Springer, 2015.

Ajt06. Miklós Ajtai. Generating random lattices according to the invariant distribution. *Draft of March*, 2006, 2006.

BCLv19. Daniel J. Bernstein, Chitchanok Chuengsatiansup, Tanja Lange, and Christine van Vredendaal. NTRU Prime. Technical report, National Institute of Standards and Technology, 2019. available at https://csrc.nist.gov/projects/post-quantum-cryptography/round-2-submissions.

BLLN13. Joppe W Bos, Kristin Lauter, Jake Loftus, and Michael Naehrig. Improved security for a ring-based fully homomorphic encryption scheme. In *IMA International Conference on Cryptography and Coding*, pages 45–64. Springer, 2013.

Che13. Yuanmi Chen. *Réduction de réseau et sécurité concrete du chiffrement completement homomorphe*. PhD thesis, Paris 7, 2013.

CJL16. Jung Hee Cheon, Jinhyuck Jeong, and Changmin Lee. An algorithm for ntru problems and cryptanalysis of the ggh multilinear map without a low-level encoding of zero. *LMS Journal of Computation and Mathematics*, 19(A):255–266, 2016.

CN11. Yuanmi Chen and Phong Q. Nguyen. BKZ 2.0: Better lattice security estimates. In *Advances in Cryptology - ASIACRYPT 2011 - 17th International Conference on the Theory and Application of Cryptology and Information Security, Seoul, South Korea, December 4-8, 2011. Proceedings*, pages 1–20, 2011.

DDLL13. Léo Ducas, Alain Durmus, Tancrède Lepoint, and Vadim Lyubashevsky. Lattice signatures and bimodal gaussians. In *Annual Cryptology Conference*, pages 40–56. Springer, 2013.

DLP14. Léo Ducas, Vadim Lyubashevsky, and Thomas Prest. Efficient identity-based encryption over ntru lattices. In *International Conference on the Theory and Application of Cryptology and Information Security*, pages 22–41. Springer, 2014.

GGH13. Sanjam Garg, Craig Gentry, and Shai Halevi. Candidate multilinear maps from ideal lattices. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 1–17. Springer, 2013.

GGH$^+$19. Nicholas Genise, Craig Gentry, Shai Halevi, Baiyu Li, and Daniele Micciancio. Homomorphic encryption for finite automata. In *International Conference on the Theory and Application of Cryptology and Information Security*, pages 473–502. Springer, 2019.

GN08. Nicolas Gama and Phong Q Nguyen. Predicting lattice reduction. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 31–51. Springer, 2008.

HPS98. J. Hoffstein, J. Pipher, and J. H. Silverman. NTRU: a ring based public key cryptosystem. In *ANTS*, 1998.

KF17. Paul Kirchner and Pierre-Alain Fouque. Revisiting lattice attacks on overstretched ntru parameters. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 3–26. Springer, 2017.

LATV12. Adriana López-Alt, Eran Tromer, and Vinod Vaikuntanathan. On-the-fly multiparty computation on the cloud via multikey fully homomorphic encryption. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 1219–1234, 2012.

LSS14. Adeline Langlois, Damien Stehlé, and Ron Steinfeld. Gghlite: More efficient multilinear maps from ideal lattices. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 239–256. Springer, 2014.

MSV09. Ivan Morel, Damien Stehlé, and Gilles Villard. H-LLL: using householder inside LLL. In *Symbolic and Algebraic Computation, International Symposium, ISSAC 2009, Seoul, Republic of Korea, July 29-31, 2009, Proceedings*, pages 271–278, 2009.

NS09. Phong Q. Nguyen and Damien Stehlé. An LLL algorithm with quadratic complexity. *SIAM J. Comput.*, 39(3):874–903, 2009.

PFH+19. Thomas Prest, Pierre-Alain Fouque, Jeffrey Hoffstein, Paul Kirchner, Vadim Lyubashevsky, Thomas Pornin, Thomas Ricosset, Gregor Seiler, William Whyte, and Zhenfei Zhang. FALCON. Technical report, National Institute of Standards and Technology, 2019. available at https://csrc.nist.gov/projects/post-quantum-cryptography/round-2-submissions.

PT08. Gábor Pataki and Mustafa Tural. On sublattice determinants in reduced bases. *arXiv preprint arXiv:0804.4014*, 2008.

Ver07. Roman Vershynin. Non-asymptotic theory of random matrices, lecture 17. https://www.math.uci.edu/~rvershyn/teaching/2006-07/280/lec17.pdf, 2007.

ZCH+19. Zhenfei Zhang, Cong Chen, Jeffrey Hoffstein, William Whyte, John M. Schanck, Andreas Hulsing, Joost Rijneveld, Peter Schwabe, and Oussama Danba. NTRUEncrypt. Technical report, National Institute of Standards and Technology, 2019. available at https://csrc.nist.gov/projects/post-quantum-cryptography/round-2-submissions.