Betweeen Codes and Lattices: Hybrid lattices and the NTWO cryptosystem
(Extended Abstract)

Cecilia Boschini, Emmanuela Orsini, Carlo Traverso

Università di Trento and IBM Research, Zürich
Department of Computer Science, University of Bristol
Dipartimento di Matematica, Università di Pisa

1 Introduction

In the past few years lattices have received considerable attention in cryptography. There are different reasons for this attention:

- The discovery of polynomial-time algorithms in the quantum computing complexity model for integer factorization and discrete logarithms [31] poses a security threat to the current public key infrastructure that relies on the hardness of these problems. Although efficient quantum computers are not currently known to exist, they might exist in the future, and information on their development might even have been withdrawn; and in any case the confidentiality of information and integrity of digital signatures should be guaranteed in view of the future technological developments. Lattice hard problems are believed to resist quantum computing attacks.
- Some hard lattice problems have been proved to allow worst-case to average case reduction. These results started with the breakthrough result of Ajtai [1], and has lead to the first of a series of proofs of security of cryptographic primitives.
- The consolidation of extremely efficient cryptographic primitives like the NTRU cryptosystem and the NTRUSign signature scheme [18,16,17], that recently have also seen the appearance of variants with proofs of security [33]. These proofs do not apply to the original NTRU system, but give an indication that the whole family is probably secure, beyond what has been currently proved.
- Lattice-based cryptography is very versatile and can be used in a variety of applications, from hash functions to cryptographic multilinear maps, from attribute-based encryption to fully homomorphic encryption, and many more ([13,10,14,29,12]).

Lattice-based cryptography is not the unique alternative to classic public-key systems. Another suitable candidate is the so called code-based cryptography. Error correcting codes are usually used to reliably transmit information over a noisy channel, but they have shown their versatility in complexity theory, and cryptography as well. We mention for example the McEliece's cryptosystem [24], based on the intractability of decoding random linear codes.

Lattices and codes have similarities and differences. They are discrete linear structures, with a nonlinear problem (discrete optimization through a distance function) that accounts for their complexity. The distance definition is where they are differentiated, and leads to different practical applications.

Where the two interact it is expected that the complexity increases. And while high complexity is bad where one aims at solving problems, it is good where designing unsolvable problems is the aim, like cryptography.

We hence define *Hybrid Lattices*, that mix Euclidean and Hamming distance. We show that these lattices model naturally a problem of polynomial algebra, that was designed to provide a "hidden ideal" cryptosystem. And our analysis shows that indeed such a cryptosystem has efficient encryption and provable security (with inefficient proof up to now, but we hope to improve it) but unfortunately the decryption is still not efficient enough to be practical.

This abstract outlines the similarities and differences between lattices (as used in cryptography) and codes, with a case study of the NTRU cryptosystem. Then we define hybrid lattices, mixing Euclidean and Hamming distance, prove how approximate optimization on them can be reduced to lattice optimization in increased dimension, and show that the cryptosystem GB-NTRU defined in [5] can be interpreted in this context, and modified in a way that might give provable security.

We show that NTWO, a modification of GB-NTRU, can resist an attack of [11] and is suitable to extend proofs of security for variants of NTRU. Its security relies on the security of a hybrid lattice. It is not efficient enough in decryption to be proposed as a realistic alternative to NTRU, but is an interesting proof-of-concept as an application of hybrid lattices.

We conclude discussing some open problems and directions for future research.

Notations

Here we collect a set of basic notations and conventions used in the paper.

- $\mathbb{Z}$ denotes the integers, $\mathbb{Z}_q$ denotes $\mathbb{Z}/(q)$.
- Tacit maps $\mathbb{Z}_q \to \mathbb{Z}$, mapping a class to a minimal weight representative. These are of course not homomorphisms, but are compatible with sum in a restricted range.
- A q-weight w on $\mathbb{Z}$ is a map $w : \mathbb{Z} \to \mathbb{R}$ that is non-negative, symmetrical, subadditive, (i.e. $w(\mathbf{x}+\mathbf{y}) \leq w(\mathbf{x}) + w(\mathbf{y})$), and $w(\mathbf{x}) = 0$ implies $\mathbf{x} \in q\mathbb{Z}^n$.
- Lattices are integer lattices, i.e. subgroups of $\mathbb{Z}^n$. A q-lattice is a lattice that contains $q\mathbb{Z}^n$.
  A q-weight defines a (q)-pseudo-distance on $\mathbb{Z}^n$, (the distance is the weight of the difference). The weight may be a norm (e.g. Euclidean) or not (e.g. Hamming). The weights are not necessarily uniform for every coordinate (this is the key to consider hybrid lattices).
- Vectors are denoted by bold lower-case letters, e.g. $\mathbf{v}, \mathbf{w}$. A polynomial f, when used as vector of coefficients, is denoted with $\mathbf{f}$ without further mention.
  Concatenation of vectors is denoted $\mathbf{v} \star \mathbf{w}$; if a, b are the weights of $\mathbf{v}, \mathbf{w}$ the weight of $\mathbf{v} \star \mathbf{w}$ is denoted by $a \star b$ (this operation depends on the context; for example, if the context of the merge is the Euclidean norm, then $a \star b = \sqrt{a^2 + b^2}$). The vector $\mathbf{e}_i$ denotes the i^{th} coordinate vector, $(0, \ldots, 0, 1, 0, \ldots, 0)$.
- Lattices are represented as matrices, the lines being a set of generators (mostly a basis). Lattices can be composed, and are shown as block matrices, blocks being lattices.

2 q-Lattices vs. q-Codes, Lee vs Hamming

Lattices used in cryptography are almost always q-lattices. Notice that a q-lattice is always of full rank, and a full-rank integer lattice is always a q-lattice, with $\det L = q$, but usually q is taken much smaller than $\det L$.

A q-code is a submodule of $\mathbb{Z}_q^n$, and there is an obvious 1-1 correspondence between q-codes and q-lattices (see e.g. [27]). The difference between the two mainly consists in the metric used: Euclidean vs. Lee or Hamming ([20]). To unify the viewpoints, a submodule of $\mathbb{Z}_q^n$ is seen as a submodule of $\mathbb{Z}^n$, and a weigh on a q-code is a q-weight on a q-lattice.

The most important problems in the algorithmic study of lattices are the Shortest Vector Problem, SVP, *i.e.* given a lattice L we are asked to

find the shortest nonzero vector of L and the Closest Vector Problem, CVP, *i.e.* given a lattice L and a target vector $\mathbf{v} \in \mathbb{Z}^n$ we are asked to find the lattice element with a minimal distance to $\mathbf{v}$. We can rephrase the CVP as the Smallest Residue Problem (SRP): given a lattice L and a target vector $\mathbf{v} \in \mathbb{Z}^n$, we have to find the smallest vector $\mathbf{v}' \in \mathbb{Z}^n$ such that $\mathbf{v} - \mathbf{v}' \in L$.

In the code setting, the corresponding problems are the Minimal Weight Codeword problem (MWC), and the Nearest Codeword Problem (NCP). With the Lee weight these are substantially the SVP and the CVP, except that the SVP for q-lattices might be a trivial solution, i.e. a $q\mathbf{e}_i$. In that case, it is tacitly assumed that the SVP means finding the shortest non-trivial solution. The CVP instead does not have exceptions.

Another small difference is that usually the Lee and Hamming distances are defined using the l_1 distance instead of the Euclidean l_2 distance (the difference is a $\sqrt{n}$ factor). This means that any algorithm for approximate CVP or SVP gives an approximate NCP or MWC, since $\sqrt{n}$ factor is usually considered small enough.

Lee and Hamming distances are instead quite different objects for distance algorithms, and using directly lattice algorithms to solve an approximate NCP in Hamming distance in a q-code is not simple. Only if it is already known that the expected solution of a Hamming MWC or NCP has very small coefficients one can use a lattice algorithm.

Note also that in Hamming MWC or NCP using l_1 or l_2 does not matter: for vectors with $\{0, 1\}$ coefficients, l_1 and l_2 norms are different, but the comparison is the same. This is however not true for Lee.

3 Four Viewpoints on NTRU

The NTRU cryptosystem ([18,16,15,17,33,22]) has become the leading candidate for a replacement of the standard cryptographic public key infrastructures based on factorization and discrete logarithm, in view of possible future technological advances that might make quantum computing at large scale realistic, and hence polynomial attacks through Shor algorithm [31] possible.

We refer to [28] for the exact definitions, and give here only an outline.

NTRU can be seen in four different ways, giving a nice illustration of the analogy between lattices and codes.

1. As a ring cryptosystem; let $G = \mathbb{Z}/(n)$ be a cyclic group, $\mathcal{A} = \mathbb{Z}_q[G] = \mathbb{Z}[X]/(q, x^n - 1)$ f, g elements of $\mathcal{A}$, invertible, sparse (with a predetermined number of monomials), with "small" coefficients (see the

details in the quoted references, but "small" mostly means in $\{0,1\}$ or $\{0,1,-1\}$), p a small element invertible in $\mathcal{A}$ (usually $p=2$, 3 or $x-1$); f should be chosen to be invertible mod p (here we use our "tacit" modular conversions).
Then $h = pgf^{-1}$ is the public key; a message $m \in \mathcal{A}$ is a sparse polynomial with "small" coefficients; to encrypt m one chooses a random sparse polynomial $r \in \mathcal{A}$ with small coefficients, and transmits $c = hr+m$. With a suitable choice of the smallness parameters, m can be recovered from c with high probability, using the fact that $hf = pg$ and reducing cf (interpreted as integer) mod p [18,17].

2. NTRU can be seen as a q-lattice cryptosystem [6]; consider the q-lattice $CS(h) \subseteq \mathcal{A} \oplus \mathcal{A}$, i.e. the submodule generated by $(h,1)$ (the Coppersmith-Shamir lattice, also called the NTRU lattice). It contains (g,f) as short vector (as shortest vector, under randomness assumptions). Moreover $(m,-r)$ is (with extremely high probability) the SRP solution for $(c,1)$. Hence both the private key and the message are protected by hard lattice problems. The lattice protecting the key and the message is the same.
This is the approach that allows most of the current cracking approaches, through standard lattice problems, hence it is the most common one ([6,11,19,9]).
3. The close similarity between q-lattices and q-codes allows to see NTRU as Lee-metric q-code. It is even the most sensible approach, except that one relies on q-lattices for attacks to q-codes anyway.
4. NTRU can also be seen as (Hamming distance) q-code cryptosystem: the minimality of (g,f) and $(m,-r)$ is true not only in the Euclidean distance, but also in the Hamming distance. (Note that this minimality of (g,f) and $(m,-r)$ is only statistically true, very special cases can be constructed in which this does not hold, even in the Euclidean distance.)

We have never seen these two last approaches mentioned, but they are both straightforward and completely useless, since they are usable neither for decryption nor for attacks, as for codes there is no tool as powerful as LLL lattice reduction and its variants. But they illustrate a point, that we use for generalizations.

4 Hybrid q-lattices

Bringing q-lattices and q-codes into a common framework allows the definition of a mixed structure, *Hybrid Lattices*.

A hybrid q-lattice is a q-lattice that uses different q-weights in different components, for example Euclidean, l_p, l_∞, Lee and Hamming (Euclidean and Hamming being the core examples); this allows a unified frame for q-lattices and q-codes.

More formally, a hybrid q-lattice is a q-lattice in $\mathbb{Z}^n$ in which is defined a q-weight as follows:

- A weight vector $\mathbf{W} = (w_i)$ is defined; each w_i is a q-weight in the ith component. If $\mathbf{x}$ is $(x_1, \ldots, x_n)$ then $\mathbf{W}(\mathbf{x})$ is $(w_1(x_1), \ldots, w_n(x_n))$. Euclidean, Lee or Hamming are the usual choices.
- A global norm $\| \ \|$ is defined (usually the Euclidean norm) .
- Given $\mathbf{v} \in \mathbb{Z}^n$, a (global) q-weight is defined as $\|\mathbf{W}(\mathbf{v})\|$

This definition allows to handle structures that mix q-lattices and q-codes. Remark that if every component is Hamming, the global distance is the square root of the usual Hamming distance. The Lee distance does not have such a simple expression, the usual Lee and Hamming distance are recovered using the l_1 norm as global norm.

We use this concept to study a cryptosystem that was first sketched in [5] as a generalization of NTRU using a private quotient of a group ring, and that can be analysed with a mix of a lattice and a (Hamming) q-code.

Roughly speaking, the private kernel is used as a source of "errors" in the key generation, and, consequently, it produces "errors" in the encryption, located in the same place. Hence the holder of the private key knows both the error locator and the short vector in the unperturbed lattice, while the attacker only knows the perturbed lattice. Details will be given later.

Approximate SVP and CVP for Hybrid Lattices. If a hybrid q-lattice has only Euclidean and Lee distance components, lattice reduction algorithms can be used to solve approximate SVP, CVP and SRP; one has just to discard, in the SVP, the possibly shorter vectors in $q\mathbb{Z}^n$, that may have weight zero.

We show here a reduction of the approximate SVP for hybrid lattices $L \subseteq \mathbb{Z}^{r+s}$ when the first r components are Euclidean and the last s are Hamming, and the projection $\mathbb{Z}^{r+s} \to \mathbb{Z}^s$ maps surjectively L to $\mathbb{Z}^s$. CVP and SRP are similar.

Generalizing a similar reduction when the projection is not surjective is a work in progress.

In this case, L can be represented as

$$\begin{bmatrix} A & 0 \\ B & I \end{bmatrix}$$

with the first column block Euclidean and the second Hamming.

We expand the lattice, depending on q, by choosing an expansion factor m and a set of *q-interpolators* $(q_1, \ldots, q_m)$. Let q_i' be the inverse of $q_i \pmod q$.

The q-lattice L of dimension $r+s$ will be replaced by a Euclidean lattice L^+ of dimension $r+ms$, in which the Hamming part is repeated m times

$$\begin{bmatrix} A & 0 & 0 & \ldots & 0 \\ B & q_1'I & 0 & \ldots & 0 \\ B & 0 & q_2'I & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ B & 0 & 0 & \ldots & q_m'I \end{bmatrix}$$

and we have a projection map $\mathbb{Z}^{r+ms} \to \mathbb{Z}^{r+s}$ sending the concatenated vector $(\mathbf{w}, \mathbf{v}_1, \ldots, \mathbf{v}_m)$ ($\mathbf{w}$ of length r, every $\mathbf{v}_i$ of length s) to $(\mathbf{w}, \sum q_i \mathbf{v}_i)$. This map sends L^+ onto L.

We prove, with an explicit construction, that every vector $\mathbf{c} = \mathbf{b} \star \mathbf{h} \in L$ ($\mathbf{b}$ of Euclidean weight w_b and $\mathbf{h}$ of Hamming weight w_h) has an inverse image $\bar{\mathbf{c}}$ in L^+ of Euclidean weight w_c such that $w_c \leq w_b \star (\xi w_h)$, with ξ that can be computed from the q_i and is independent of the dimensions r and s. (The proof is omitted in this abstract).

Given the q_is the value of ξ can be computed for $(r,s) = (0,1)$, and also the expected average value ξ' can be easily computed (at least for small q and very small m, as they are usually). Hence given q one can either find an optimal q_i or check the quality of a heuristic choice (q_1 can always be put equal to 1). This allows to evaluate the tradeoff between increased lattice expansion and increased weight expansion.

5 Variants of NTRU

In this section we describe different variants of NTRU that are easy to design using a different ring $\mathcal{A}$. In the commutative case, $V(\mathcal{A})$ will denote $\mathsf{Spec}(\mathcal{A})$, the associated algebraic variety. We will use freely the algebraic geometry language (points, support, etc.).

Group variant. One can use a different group, for example a product of cyclic groups ($\mathcal{A}$ in this case is a quotient of a multivariate polynomial ring). This approach is not convenient, since in this way some weaknesses are introduced, see the Gentry attack below. Other (non-commutative) groups have been proposed, without much success up to now.

These more general NTRU variants, have a *dimension* (as free $\mathbb{Z}_q$ module) and an *order* (the maximum order of the elements of the group).

NTRU+ and NTRU$_s$. The NTRU cyclic group ring algebra is just $\mathbb{Z}_q[X]/(x^n - 1)$. One might consider other quotients of $\mathbb{Z}_q[X]/(f)$, but to generalize NTRU it is necessary that multiplication by x is an isometry. The only other possible choice is hence $x^n + 1$. We call this variant NTRU+.

If n is odd, $\mathbb{Z}_q[X]/(x^n+1)$ may be mapped isomorphically to $\mathbb{Z}_q[X]/(x^n-1)$ sending x to $-x$; if n is the product of an odd number and a power of 2 it might be decomposed in two parts, one corresponding to the odd part and the other corresponding to the power of 2, so it is really something different only for n power of 2. Notice that since $x^{2^{r+1}}-1 = (x^{2^r}-1)(x^{2^r}+1)$, the NTRU ring for $n = 2^{r+1}$ can be split as direct sum of two rings: one is an NTRU ring, the other is an NTRU+ ring, both for $n = 2^r$. Considering this, NTRU+ in dimension 2^r has order $n = 2^{r+1}$.

This variant has been used by Stehlé and Steinfeld in [33]. They slightly modify parameters, key generation and encryption to obtain a version that allows a proof of security through reduction of worst-case to average-case complexity. More precisely

- Use NTRU+ in dimension $n = 2^r$, with q such that $2n$ divides $q - 1$.
- The encryption is made as $c = m+pe+hr$ instead of $c = m+hr$, with e, r random (from appropriate distributions). The additional term e is required to achieve IND-CPA security.
- Special sampling rules are used to satisfy the results concerning worst-case to average-case reduction (this is a moving target, and new results seem to imply that these rules might be relaxed).

The decryption through the private key remains unchanged. We call this version NTRU$_s$.

Parameters determined using the performance of the current state of the art of lattice reduction and CVP algorithms has allowed an implementation that is reasonably efficient, although not enough to be considered practical [4]. Recent results however might allow to improve some parameters, resulting in increased efficiency.

6 NTWO, a Hybrid Lattice Cryptosystem

The cryptosystem that we now call NTWO has been first sketched in [5] with the name GB-NTRU, and defined as a ring cryptosystem. It has been later reformulated as a hybrid lattice cryptosystem, based on bivariate polynomial NTRU, and renamed NTWO, but never published (it has been exposed at some poster sessions).

We reformulate it now, changing the encryption adding an extra pseudo-random term, considering as basis the univariate or bivariate NTRU rings, or NTRU+ rings. To simplify the discussion, we suppose $p = 2$. We will use GB-NTRU to refer to the version of [5] and reserve NTWO for the current, modified formulation.

Originally, the NTWO name was adopted since we mainly regarded the bivariate version, (the main reason is to allow smaller q with respect to the dimension). We keep it also for the univariate case, that in the current discussion is important too. In the bivariate case, we suppose for simplicity that both variables have the same minimal polynomial, although the general case has been tested too.

NTWO as ring cryptosystem

The basic idea, in the ring setting, is to have two rings, the public one, $\mathcal{A}$, being the NTRU ring (univariate or bivariate, the NTRU or the NTRU+ ring) and the private one, $\bar{\mathcal{A}} = \mathcal{A}/\mathcal{I}$ being a quotient of $\mathcal{A}$ modulo a small private ideal $\mathcal{I}$, "small" being its dimension as $\mathbb{Z}_q$-vector space, or equivalently its "support", i.e. the complementary of the zero locus (on the algebraic closure).

The modulus q is a prime number (it is possible to generalize to q composite, but this will just complicate the geometry and obscure the ideas) and we further assume that all the points of $V(\mathcal{A})$ are $\mathbb{Z}_q$ rational. This means that the order of $\mathcal{A}$ should divide $q - 1$.

The ideal $\mathcal{I}$ may be any ideal, but it is empirically necessary that the support does not contain any point whose coordinates are in the set $\{1, -1\}$. This condition is a technical need, because the monic univariate polynomials with support in $\{1, -1\}$ have coefficients in $\{1, -1\}$, and this makes decryption harder and key attacks easier.

The private ring is used to prepare the public key and to decrypt; the public ring is used for encryption.

As we did for NTRU, we do not specify here what is "small", but of course exact specifications are needed, how f, g, r, I, e, etc. are chosen randomly, once $\mathcal{A}$ and p are fixed.

Key generation. The public key is generated as in NTRU as $\bar{h} = pg/f \in \bar{\mathcal{A}}$, and is lifted to $h \in \mathcal{A}$ adding to $\bar{h}$ a random element $\alpha \in \mathcal{I}$ (but with the condition that neither h nor α has a zero. The secret key consists of (f, g, α). Note that α can be deduced from f, g and the public key h, and $\mathcal{I}$ is generated by α.

Encryption. To encrypt, we use a modified form of NTRU+ encryption. More precisely, given a message $m \in \mathcal{M}$, we compute a random $r \in \mathcal{A}$ and a pseudo-random e (it should be computed via a secure hash of m, so that it can be recovered knowing m but not conversely). Then the cryptogram will be $hr + m + pe$ (remark that this is similar, but not identical to the e used in NTRU$_s$. Smallness conditions similar to NTRU (or NTRU$_s$) should hold, to allow decryption. We don't define them here, but are of course essential.

Decryption. Decryption of a cryptogram c is done in two steps. The first step, computes cf. The result of the first step is not $fm+pgr$ as in NTRU, but it is $fm + pgr + pfe + \alpha'$, where $\alpha' = fr\alpha$ is a (random) element of the private ideal $\mathcal{I}$ (a random "error" whose location is known). This "error" is removed through a solution of an approximate CVP (see [3]) in the private q-lattice $\mathcal{I}$ (it is assumed that the key creation parameters are such that $fm + pgr + pfe$ is statistically short). Then m is recovered as in NTRU, through removal of multiples of p hoping (i.e. proving that the opposite event is extremely unlikely) that no carry has messed with them. Recovering m allows to find e (re-hashing m) and r. Checking that c has been produced conforming to the specifications allows to detect possible attacks based on a CCA.

Decryption may fail for two reasons: either the approximate CVP algorithm fails to find a sufficiently small vector, or the vector found is smaller than $fm + pgr + pfe$. This for example would be the case if the support of $\mathcal{I}$ contained $(\pm 1, \pm 1)$, a possibility that we have excluded. Moreover the decryption may fail for the same reasons why NTRU may fail, i.e. the heuristics extracting fm mod p from $fm + pgr + pfe \in \mathcal{A}$ fails.

NTWO as (Hybrid) Lattice Cryptosystem

Having two different rings, we have two different lattices, both public. One describes the message encryption, and the other describes the key generation. NTRU has just one, that is used for both, and this is a kind of weakness: making a key that is at the same time robust for key attacks

and message attacks, and still allowing decryption, is a difficult balance for NTRU.

The public key h can be used to build a Coppersmith-Shamir lattice, like the NTRU lattice, and this can be used for message attacks. A different one has to be used for key attacks.

Message attack Encryption is more or less equivalent to NTRU encryption. The addition of the extra e has as only consequence that computing a SRP for the cryptogram $m + pe + hr$ with respect to the public CS lattice is expected to give $(m + pe, r)$. From this m is easily recovered. The task is however harder than the key attack for NTRU, since $m + pe$ is larger than m.

Key Attacks A SVP for the public NTRU lattice does not give anything usable to recover the private key, consisting of f, g, α. Recall, α is a polynomial whose support is the support of the private ideal, and the attacker only knows that its support is small. This means that, denoting by λ_P the polynomial whose value is 1 in a point P of $V(\mathcal{A})$ and 0 elsewhere, we have $\alpha = \sum_P a_P \lambda_P$ and the Hamming weight of the vector $\mathbf{a} = (a_P)$ is the cardinality of the support of I. The λ_p are called Lagrange Interpolators, and they form a $\mathbb{Z}_q$-basis of the ring $\mathcal{A}$.

Hence to recover the private key, one has to recover (f, g, α), that we represent as a vector $\mathbf{f} \star \mathbf{g} \star \mathbf{a}$; so we need an extension of the CS lattice, the LCS lattice (the Lagrange-Coppersmith-Shamir lattice) that we define here. It is a hybrid lattice in $\mathcal{A} \oplus \mathcal{A} \oplus \mathcal{A}$, in which the first two components have Euclidean weight and the third has Hamming weight; it is defined with blocks as follows:

$$\begin{bmatrix} qI & 0 & 0 \\ H & I & 0 \\ L & 0 & I \end{bmatrix}$$

in which H represents the monomial multiples of h, and L has the rows representing the λ_P as sum of monomials. This is not only 50% larger than the CS lattice, but being a hybrid lattice with a Hamming block special algorithms should be used.

Key Attacks from Partial or Special Information. We omit the proof of the following results, that are relatively simple:

- Knowing f and g allows to find α computing $h - \bar{h}$, and $\mathcal{I}$ as the locus where gf^{-1} differs from h.

- Knowing $\mathcal{I}$, recovering f and g still requires to solve a SVP in a (standard) q-lattice, obtained from the LCS lattice preserving just the rows corresponding to λ_P where P is in the support of the ideal; hence a problem harder than an NTRU key recovery problem.
- Knowing $\mathcal{I}$ and f (or g) still requires a SVP on a q-lattice with halved dimension with respect to the CS lattice.
- In the case that α has been created as $\sum a_p \lambda_p$ and the a_p are small, recovering the private key means solving a SRP in LCS with Euclidean norm.

Chosen Ciphertext Attacks. The decryption procedure requires that after finding m one recovers e (hashing m) and r (by difference and division by the public key ph). If r does not conform to the specifications of the protocol, this means that extra errors have been introduced, and the message is probably part of a side-channel attck. Hence the decryption should be considered a failure.

If the final recovery of e and r is skipped, an easy CCA can be successful and decides if a point is in the private ideal: adding $c\lambda_P$ to a valid cryptogram, produces another cryptogram whose decryption gives the same message if and only if the interpolator is in the private ideal (if not, the result will be almost surely a decrypting failure); however, in this case the random element r', deduced from the manipulated cryptogram, also has a multiple of λ_P added, hence cannot be small (the only interpolator with almost constant coefficients is the interpolator of one of the points that we have excluded from the private ideal).

This is the main reason why e in the encryption protocol has to be deduced from the message m, instead of being determined randomly.

7 The Gentry Attack

NTRU can be subject to a class of homomorphism attacks, whose prototype is discussed in [11]. We discuss the Gentry attack for NTRU, NTRU+ and NTWO in the same context, just using the fact that they are based on a quotient ring $\mathcal{A}$ of a group ring, that identifies a lattice L.

The attack can be mounted whenever there is a surjective homomorphism from an NTRU-like ring to an NTRU-like ring, that is "locally" an isometry, i.e. whenever the image of an element of weight one is an element of weight one. These homomorphisms should not be injective, since in that case it will be an isomorphism and an isometry with the image. In the case of "classic" NTRU such homomorphisms exist when

the dimension n is composite, since if $n = cd$ then $x^c - 1$ divides $x^n - 1$; and this is the case considered in [11]. This falls into two subcases, the first one is $n = cd$ with c, d coprime, and the second one is n power of a prime number – notably 2 being the only reasonable option. In either case, if a factor is small, (either d or 2) one can attack both the key and the message through a direct attack to the quotient, that is of smaller dimension, and then lifting the result to the full lattice using different techniques, as sketched in [11].

The attack may fail either by failing the SVP (for the key attack) or the CVP (for the message attack) in the quotient, or because the result is not the image of the result of the corresponding original problem. The first issue can be solved using stronger algorithms (recall the dimension of the quotient lattice is sufficiently small), but the second is much harder to deal with. The paper reports a high rate of total success, i.e. recovering either the key or the message, for n up to 256.

In the case of NTRU+, if n is odd, nothing changes, since the map $x \to -x$ maps the NTRU+ ring to the NTRU ring. If the dimension is 2^n there is no suitable homomorphism.

In the bivariate case, the attacker has more powerful weapons, since there are many suitable homomorphisms from the bivariate NTRU ring $\mathbb{Z}[x, y]/(x^n - 1, y^n - 1)$ (this is the only case that is worth considering) to a univariate NTRU ring $\mathbb{Z}[t]/(t^n - 1)$; the suitable maps are all the maps $\phi_{r,s} : x \to t^r,\ y \to t^s$ but the pairs (r, s) and (ar, as) have the same kernel, hence we have $n + 1$ homomorphisms. Although even a success with one image (not considering the lifting) proves the setting insecure, detecting correctly every one of the different $n+1$ images allows a recovery of the information through linear algebra (the resulting system is even overdeterminate).

We tested several cases with $p = 2$ and $n = 13,\ 17,\ 19,\ 23$, that are the reasonable parameters to test (giving global dimension of the NTRU ring 169, 289, 361, 529), and we had a high rate of complete success (correct identification of all the quotients), just with LLL reduction for SVP and the algorithms of [3] for CVP, and almost no example of complete failure. So bivariate NTRU has to be considered completely insecure.

The scenario changes with GB-NTRU, (and much more with NTWO). Now the key attack and the message attacks require to consider different lattices.

We first analyse the message attack, with the CS lattice. Here experimentally the attack mostly succeeds for GB-NTRU, that is similar to NTRU in the encryption: we have large rate of full successes, and a

general recovery of at least one, but often several, images of the message. So GB-NTRU has to be considered broken. Moreover (with $p = 2$) it is very easy to deduce that the attack succeeded, since in this case the image of the message is a vector with non-negative coordinates (this being invariably the sign of success).

This is radically different for NTWO: the addition of e in the encryption makes the possible successes unrecognizable, but with a suitable choice of the parameters for the variability of e one always has failure because the image of the message is not the shortest residue with respect to the quotient lattice.

For the key attack, there is no difference between GB-NTRU and NTWO; we have to map the LCS lattice, from bivariate to univariate. Now the problem is more difficult, since the CS part can pass to the quotient, but this is not true for the Lagrange part. A λ_P mapped to univariate is not a univariate interpolator, but is a product of two univariate interpolators. Hence, while the CS part can be reduced from dimension n^2 to dimension n, the L part is reduced from n^2 to approximately $n^2/2$, hence it remains too hard.

We can hence conjecture that (bivariate) NTWO is practically secure: it is harder than bivariate NTRU. Extrapolating, we may guess that univariate NTWO is more secure that NTRU. We give further hints in the next section.

8 Lattices and Security

The security of a cryptosystem relies on two aspects. The theoretical aspect is a proof of a theorem (often based just on a conjecture) that some problem is (asymtoptically) hard in the average case of a suitably chosen subset of problems. The practical aspect relies on the current state of the art, algorithmic and technical, that may give an estimate of the hidden constants of the asymptotic formulas, in terms of time needed to solve a hard problem. Looking at the theoretical aspect more in detail, one wants to design systems that are "provable secure", i.e. that admit a reduction to some "assumed hard" problems. In particular this means that even a provable secure system could be insecure, for example if the underlying hardness assumption turns out to be false.

Lattice cryptosystems are especially prized since they have been proved secure, i.e. hard to break on the *average*, based on the assumption that the underlying lattice problems are hard in the *worst-case*.[1]

These reduction theorems mostly rely on two hard problems: the SIS problem [1] and the LWE problem [29]. The former one is the problem of finding short (nonzero) vectors in random lattices; the latter asks to recover a secret $\mathbf{s}$ from a sequence of noisy random linear equations in $\mathbf{s}$. Both of them admit generalizations for (cyclotomic) rings [23,25] and, more generally, for large enough abelian group rings ([8]).

To deduce that an explicitly chosen parameter set gives rise to a practically secure cryptosystem is however necessary to have heuristic estimates of the difficulty of solving a worst case problem in a given finite set of problems. This difficulty can only be estimated by an analysis of the performance of the state-of-the-art algorithms, extrapolating their behaviour from the feasible problems to the unfeasible ones. The reduction of worst-case to average-case complexity guarantees the likelihood that a computational experiment (performed on a random choice) gives results applicable to any other random choice.

The weak point is that it is impossible to prove that the performance of an algorithm is optimal, since it is always possible that an algorithmic or a technological advancement or a new theorem allows an unsuspected improvement. This happened for example with the introduction of the LLL algorithm of [21] and related variants [30] and optimized implementation [32], and might happen with quantum computing [31].

9 Security of NTWO

While a proof of security of NTWO (univariate and bivariate, mainstream variant or NTRU+ variant) should rely on a formal reduction of worst case to average case, and on a practical analysis of the known algorithms and their efficiency, this is still a work in progress. We list here a few facts that show that some specially selected keys coupled with the disclosure of some information provides security greater than the corresponding NTRU or NTRU+ key. This is not satisfactory since in this way we obtain results much weaker than what we expect to obtain in a near future, but it is all that we have now.

[1] Hardness results in cryptology are often called "Assumptions"; a proof of reduction is a theorem, but the conclusion that a problem is hard is just an assumption, based on the (widely shared) assumption that another problem is hard. And at best this is reduced to the basic assumption, $P \neq NP$.

- The security of NTRU$_s$ relies on two assumptions, namely the Ring Learning With Errors, RLWE, and the Decisional Small Polynomial Ratio, DSPR [33], [22]. The main difference with NTRU is in the key generation, in that f and g are sampled from a discrete Guassian distribution with a large enough standard deviation. In this way the DSPR ensures that the public key h is statistically close to uniform. The encryption is identical of that of NTWO, except that e is again sampled from a Gaussian distribution, and the indistinguishability of the ciphertext from a random element in $\mathcal{A}$ is deduced from RLWE.
- At a high level, we would like to formally prove the security of NTWO relying on two assumptions. We can call the first one noisy-DSPR. More precisely we relax the hypothesis of DSPR and consider f, g as random small polynomials in $\mathcal{A}$, instead of being generated from a Gaussian distribution, as in the "classic" NTRU, and then we add to $\bar{h} = g/f$ a small (in the Hamming distance) error α . Although there is not a formal proof that $\bar{h}$ generated as the quotient g/f is indistinguishable from random in $\mathcal{A}$, this problem remains unsettled (in both of its versions, search and decision) after almost twenty years of cryptoanalysis. So even disclosing the private ideal, we still need to solve the original NTRU assumption. The second assumption should be a variant of RLWE, that states that the distribution $(h, hr + pe)$ is indistinguishable from uniform, where h is uniform in $\mathcal{A}$ and r, e are small polynomials in $\mathcal{A}$. For details about the right choice of e and r see [2,26,7].

- Disclosing the private ideal, one still has to identify the coefficients of $\alpha = \sum a_P \lambda_P$, with P in the support of the ideal. This is a SVP in a sublattice of the LCS lattice (limited to the rows with the λ_P in use). The metric is Euclidean, with the columns corresponding to the 1 in the λ_P rows, with weight very small (for example, $1/q$)

Implementation, Experiments and Practical Security of NTWO

We have an experimental implementation of NTWO, that we used for tests. The weak point of the cryptosystem is that decryption is slow, and some ideals fail with non-negligeable probability to allow decryption. This may happen for certain choices of parameters, but often with the same choice of parameters it seems that some ideals perform well, and other perform badly. We have been unable up to now to disclose a pattern. It seems that the choice of f, g and α once the ideal points are fixed do not have a similar influence.

We plan to produce a technical report on the experiments, but up to now they have been useful mainly to explore; for example, the addition of a pseudo-randomization in encryption has been suggested by experiments with the bivariate Gentry attack, and formalized in a way similar to NTWO_s just to try to match a provable security proof.

We also have an implementation of hybrid lattices SVP, that works very well to crack the NTRU key in very small cases (5×5 being our record for now).

If the ideal is too small doing an exhaustive search of the ideal (attempting to crack the key considering every possible ideal of the expected size) might be possible, and for very low dimension (up to 11×11 and 13×13) tuning the cardinality of the support of $\mathcal{I}$ to allow decryption and at the same time discourage an exhaustive search for the ideal might be challenging, but with 17×17 there is usually no problem. But the choice of q too has a rôle.

Direct message attacks seem to be very hard (as expected) in these cases.

Univariate NTWO+ is very promising too. Dimension 128 and $q = 257$ deserves to be explored well.

We have not yet experimented the bivariate NTWO+ in even dimension. Interesting cases to explore would be 16×16 (possible values of $q = 97,\ 193,\ 257,\ 353,\ 449,\ 577,\ 641, \dots$) or 32×32, ($q = 193,\ 257,\ 449,\ 577,\ 641,\ 769,\ 1153,\ 1217 \dots$. One might also explore the mixed cases 16×32, 16×64 and 32×64 ($q = (257,\ 641,\ 769,\ 1153,\ 1409)$. The case 64×64 is probably too large to be significant, the smaller the dimension the easier is to see the problems). The advantage of bivariate here too is the larger choice of q with respect to the same dimension in univariate.

10 Work in Progress and Open Problems

Worst-Case to Generic Reduction for Hybrid Lattices

This would be the ideal completion of the proposal of hybrid lattices; the analysis of injected errors seems to fit very well in the LWE setting, and the analysis of NTRU_s relies on a RLWE. He4nce extending it should be possible.

Randomize m in Encryption

In the NTWO encryption $m + pe + hr\ e$ is taken pseudo-random being obtained as a hash of m. This is unsatisfactory, since the same message

will always use the same e. Even if using a cryptographic hash might ensure pseudo-randomicity, and the randomization given by r might be enough, a proof of security might become more difficult.

A possible modification might be to randomize the message, using random errors and an error-correcting code. So one migh have a public ECC, encode the message m, add a fixed number of correctable random errors obtaining m', compute e from the modified message. Although probably this might either not be necessary, or be insufficient, it is a possibility to consider, and exhibits another mixed feature between lattice cryptography and codes.

Reduction of q-codes

To decrypt NTWO the hard part is the SRP on the private lattice. This currently is done reducing the corresponding lattice. This means reducing a lattice of high dimension associated to a code of low dimension. This is usually fast enough, but an algorithm performing a reduction directly on the low-dimensional code should be much better.

Being able to work in a low-dimensional setting (the code, not the space in which it is embedded) should allow to use *ad hoc* algorithms for the CVP.

Impact of Private Ideals on Decryption

The interpolator of $(1, 1)$ is the polynomial $\sum x^i y^j$ and its presence in the ideal support usually makes the heuristics for decryption fail. This has been already remarked, and taken into account.

The existence of a point $(1, a)$ or $(b, 1)$ in the support seems to produce harder private lattices; especially if there are several ones. It might be useful to avoid these points too. This deserves to be further investigated; for now we don't have collected any statistical evidence. Collecting a large set of examples of ideals, and discovering relations between the ideal quality and its geometric properties could be interesting.

NTWO signature

NTRU has a companion NTRUSign signature algorithm. NTRU_s too has a signature algorithm. A NTWOSign algorithm seems to be harder, but the issue has not been investigated sufficiently, partly due the the several signature variants that have been proposed.

We would like also to investigate other cryptographic applications of hybrid lattices.

Conclusions

We have defined hybrid lattices, mixing different weights, and shown that their use might give new hard problems and might be used to modify existing lattice cryptograpic protocols improving their security without too much penalizing the performance, especially in encryption. This is opens a new research area in an already trendy topic.

References

1. M. Ajtai. Generating hard instances of lattice problems (extended abstract). In *Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing, Philadelphia, Pennsylvania, USA, May 22-24, 1996*, pages 99–108, 1996.
2. S. Arora and R. Ge. New algorithms for learning in presence of errors. In *Automata, Languages and Programming - 38th International Colloquium, ICALP 2011, Zurich, Switzerland, July 4-8, 2011, Proceedings, Part I*, pages 403–415, 2011.
3. L. Babai. On lovász' lattice reduction and the nearest lattice point problem. *Combinatorica*, 6(1):1–13, 1986.
4. D. Cabarcas, P. Weiden, and J. Buchmann. On the efficiency of provably secure NTRU. In *Post-Quantum Cryptography - 6th International Workshop, PQCrypto 2014, Waterloo, ON, Canada, October 1-3, 2014. Proceedings*, pages 22–39, 2014.
5. M. Caboara, F. Caruso, and C. Traverso. Gröbner bases for public key cryptography. In *Symbolic and Algebraic Computation, International Symposium, ISSAC 2008, Linz/Hagenberg, Austria, July 20-23, 2008, Proceedings*, pages 315–324, 2008.
6. D. Coppersmith and A. Shamir. Lattice attacks on NTRU. In *Advances in Cryptology - EUROCRYPT '97, International Conference on the Theory and Application of Cryptographic Techniques, Konstanz, Germany, May 11-15, 1997, Proceeding*, pages 52–61, 1997.
7. N. Döttling and J. Müller-Quade. Lossy codes and a new variant of the learning-with-errors problem. In *Advances in Cryptology - EUROCRYPT 2013, 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Athens, Greece, May 26-30, 2013. Proceedings*, pages 18–34, 2013.
8. N. Gama, M. Izabachène, P. Q. Nguyen, and X. Xie. Structural lattice reduction: Generalized worst-case to average-case reductions. *IACR Cryptology ePrint Archive*, 2014:283, 2014.
9. N. Gama and P. Q. Nguyen. New chosen-ciphertext attacks on NTRU. In *Public Key Cryptography - PKC 2007, 10th International Conference on Practice and Theory in Public-Key Cryptography, Beijing, China, April 16-20, 2007, Proceedings*, pages 89–106, 2007.
10. S. Garg, C. Gentry, and S. Halevi. Candidate multilinear maps from ideal lattices. In *Advances in Cryptology - EUROCRYPT 2013, 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Athens, Greece, May 26-30, 2013. Proceedings*, pages 1–17, 2013.
11. C. Gentry. Key recovery and message attacks on ntru-composite. In *Advances in Cryptology - EUROCRYPT 2001, International Conference on the Theory and Application of Cryptographic Techniques, Innsbruck, Austria, May 6-10, 2001, Proceeding*, pages 182–194, 2001.

12. C. Gentry. Fully homomorphic encryption using ideal lattices. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*, pages 169–178, 2009.
13. C. Gentry, C. Peikert, and V. Vaikuntanathan. Trapdoors for hard lattices and new cryptographic constructions. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008*, pages 197–206, 2008.
14. S. Gorbunov, V. Vaikuntanathan, and H. Wee. Attribute-based encryption for circuits. In *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 545–554, 2013.
15. P. S. Hirschhorn, J. Hoffstein, N. Howgrave-Graham, and W. Whyte. Choosing ntruencrypt parameters in light of combined lattice reduction and MITM approaches. In *Applied Cryptography and Network Security, 7th International Conference, ACNS 2009, Paris-Rocquencourt, France, June 2-5, 2009. Proceedings*, pages 437–455, 2009.
16. J. Hoffstein, N. Howgrave-Graham, J. Pipher, J. H. Silverman, and W. Whyte. NTRUSIGN: digital signatures using the NTRU lattice. In *Topics in Cryptology - CT-RSA 2003, The Cryptographers' Track at the RSA Conference 2003, San Francisco, CA, USA, April 13-17, 2003, Proceedings*, pages 122–140, 2003.
17. J. Hoffstein, N. Howgrave-Graham, J. Pipher, and W. Whyte. Practical lattice-based cryptography: NTRUEncrypt and NTRUSign. In *The LLL Algorithm - Survey and Applications*, pages 349–390. Springer, 2010.
18. J. Hoffstein, J. Pipher, and J. Silverman. Ntru: A ring-based public key cryptosystem. In J. Buhler, editor, *Algorithmic Number Theory*, volume 1423 of *Lecture Notes in Computer Science*, pages 267–288. 1998.
19. N. Howgrave-Graham. A hybrid lattice-reduction and meet-in-the-middle attack against NTRU. In *Advances in Cryptology - CRYPTO 2007, 27th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2007, Proceedings*, pages 150–169, 2007.
20. C. Lee. Some properties of nonbinary error-correcting codes. *Information Theory, IRE Transactions on*, 4(2):77–82, June 1958.
21. A. Lenstra, J. Lenstra, H.W., and L. Lovász. Factoring polynomials with rational coefficients. *Mathematische Annalen*, 261(4):515–534, 1982.
22. A. López-Alt, E. Tromer, and V. Vaikuntanathan. On-the-fly multiparty computation on the cloud via multikey fully homomorphic encryption. In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 1219–1234, 2012.
23. V. Lyubashevsky, C. Peikert, and O. Regev. On ideal lattices and learning with errors over rings. In *Advances in Cryptology - EUROCRYPT 2010, 29th Annual International Conference on the Theory and Applications of Cryptographic Techniques, French Riviera, May 30 - June 3, 2010. Proceedings*, pages 1–23, 2010.
24. R. J. McEliece. A public-key cryptosystem based on algebraic coding theory. *DSN Progress Report*, (42-44: 114), 1978.
25. D. Micciancio. Improved cryptographic hash functions with worst-case/average-case connection. In *Proceedings on 34th Annual ACM Symposium on Theory of Computing, May 19-21, 2002, Montréal, Québec, Canada*, pages 609–618, 2002.
26. D. Micciancio and C. Peikert. Hardness of SIS and LWE with small parameters. In *Advances in Cryptology - CRYPTO 2013 - 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part I*, pages 21–39, 2013.

27. D. Micciancio and O. Regev. Lattice-based cryptography. In D. J. Bernstein and J. Buchmann, editors, *Post-quantum Cryptography*. Springer, 2008.
28. P1363.1-2008. Standard specification for public key cryptographic techniques based on hard problems over lattices. *IEEE*, 2009.
29. O. Regev. On lattices, learning with errors, random linear codes, and cryptography. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing, Baltimore, MD, USA, May 22-24, 2005*, pages 84–93, 2005.
30. C. P. Schnorr and M. Euchner. Lattice basis reduction: Improved practical algorithms and solving subset sum problems. In *Math. Programming*, pages 181–191, 1993.
31. P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.*, 26(5):1484–1509, 1997.
32. V. Shoup. Ntl: A library for doing number theory.
33. D. Stehlé and R. Steinfeld. Making NTRU as secure as worst-case problems over ideal lattices. In *Advances in Cryptology - EUROCRYPT 2011 - 30th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tallinn, Estonia, May 15-19, 2011. Proceedings*, pages 27–47, 2011.