Des. Codes Cryptogr. (2010) 56:65–78
DOI 10.1007/s10623-009-9342-7

NTRU over rings beyond $\mathbb{Z}$

Monica Nevins · Camelia KarimianPour · Ali Miri

Received: 5 March 2009 / Revised: 8 September 2009 / Accepted: 22 October 2009 /
Published online: 10 November 2009

Abstract The NTRU cryptosystem is constructed on the base ring $\mathbb{Z}$. We give suitability conditions on rings to serve as alternate base rings. We present an example of an NTRU-like cryptosystem based on the Eisenstein integers $\mathbb{Z}[\zeta_3]$, which has a denser lattice structure than $\mathbb{Z}$ for the same dimension, and which furthermore presents a more difficult lattice problem for lattice attacks, for the same level of decryption failure security.

Keywords Public-key cryptography · Lattice · NTRU

Mathematics Subject Classification (2000) Primary 11T71 · Secondary 13G

1 Introduction

NTRU is a public-key cryptosystem introduced by Hoffstein et al. in 1996 [7]. Its security is conjectured to rely on the hardness of the shortest and closest vector problems in a lattice. NTRU is based on the ring of polynomials with coefficients in $A = \mathbb{Z}$. While demonstrably efficient and practical, with less overhead and smaller key sizes than other public key cryptosystems such as RSA and ECC, its decryption algorithm has a nonzero probability of failure. This failure is related to properties of the base ring A.

Communicated by S. Galbraith.

M. Nevins (✉) · C. KarimianPour
Department of Mathematics and Statistics, University of Ottawa, Ottawa, Canada
e-mail: mnevins@uottawa.ca

C. KarimianPour
e-mail: ckari099@uottawa.ca

A. Miri
School of Information Technology and Engineering (SITE), and Department of Mathematics and Statistics, University of Ottawa, Ottawa, Canada
e-mail: samiri@site.uottawa.ca

In 2002, Gaborit et al. proposed an analogue of NTRU, called CTRU, in which the base ring A is replaced by $\mathbb{F}_2[x]$, the ring of polynomials over a binary field. They showed that CTRU had no decryption failures; however, in 2005 Kouzmenko [11] showed that CTRU was open to a deterministic polynomial-time attack. Kouzmenko, in turn, discussed possible conditions on alternatives to $A = \mathbb{Z}$ or $\mathbb{F}_2[x]$, and presented the example of GNTRU, based on $A = \mathbb{Z}[i]$, the Gaussian integers. Other works have followed, including MaTRU [2] by Coglianese and Goi in 2005, based on matrix rings.

In this paper, we extend Kouzmenko's ideas and develop criteria for the suitability of a (commutative) ring A for use as an NTRU base ring. These results are presented in Theorem 4.1, Corollary 4.2 and Theorem 4.3. We discuss in detail the role of having a unique remainder in the Euclidean function. To illustrate our theorems, we explicitly develop the groundwork for the case of the Eisenstein integers $A = \mathbb{Z}[\zeta_3]$, which we call ETRU. In particular, in Theorem 5.2, we demonstrate the connection between the decryption algorithm and the lattice structure of the base ring, and then use a probabilistic model to argue, in Proposition 6.1, that ETRU will have a comparable level of decryption failure security for same-sized parameters of NTRU or GNTRU, while having a higher level of combinatorial security. Finally, we present the ETRU lattice in Proposition 7.1 following [4], in which solving the shortest vector problem implies recovering the private keys. In contrast with NTRU and all previous generalizations, this lattice is non-rational. Furthermore (Corollary 7.2) ETRU has a higher lattice constant than do NTRU or GNTRU at the same dimension, implying higher lattice security.

This paper is organized as follows. We briefly summarize the NTRU cryptosystem in Sect. 2. Some necessary mathematical background is presented in Sect. 3. In Sect. 4, we present our main theorems regarding alternative base rings for NTRU, and in Sect. 5 we show that $\mathbb{Z}[\zeta_3]$ with its dense lattice structure can be used as a base ring for NTRU. The decryption failure security of ETRU is discussed in Sect. 6. The ETRU lattice problem is defined and discussed in Sect. 7. In Sect. 8, we offer our conclusions as well as suggestions for future work.

2 The NTRU cryptosystem

The material in this section is summarized from [7].

The NTRU cryptosystem depends on a set of six positive integer parameters (N, p, q, d_f, d_g, d), where p and q are relatively prime, q is considerably bigger than p and d, d_g and d_f are positive integers less than $N/2$. Given such a set of parameters, one sets $R = \mathbb{Z}[X]/(X^N - 1)$ and chooses subsets L_f, L_g, L_ϕ and L_m of R (related to the values of d_f, d_g and d). Let $R_q = (\mathbb{Z}/q\mathbb{Z})[X]/(X^N - 1)$ and similarly for R_p. Given $\overline{a} \in R_q$, we identify it with the representative $a \in R$ which has coefficients in the interval $[-q/2, q/2)$.

To generate the keys, choose f and g randomly from L_f and L_g respectively, under the condition that the images of f in each of R_q and R_p should be invertible; denote these inverses by F_q and F_p, respectively. The private key is f; the public key is $\overline{h} = gF_q \in R_q$.

To encrypt a message $m \in L_m$, choose $\phi \in L_\phi$ at random and compute the ciphertext $\overline{c} = p\phi\overline{h} + m \in R_q$. To decrypt $\overline{c}$, make the mod q calculation:

$$\overline{a} = f\overline{c} = f(p\phi g F_q + m) = p\phi g + fm \in R_q$$

and identify the answer $\overline{a}$ with the unique polynomial $a \in R$ with coefficients in the interval $[-q/2, q/2)$ whose reduction mod q is $\overline{a}$. If it is true that $a = p\phi g + fm \in R$, that is, that the result of the modular calculations coincides with the expected non-modular result, then we obtain m from the mod p calculation:

$$F_p a = pF_p \phi g + F_p f m = m \in R_p.$$

Otherwise, we say that there was a *decryption failure*. A necessary condition for successful decryption is that the polynomial $a = \sum a_i x^i$ satisfies $|a|_\infty := \max a_i - \min a_i < q$. The parameters for NTRU are chosen so that the probability of this type of decryption failure is minimal, by ensuring that the standard deviation of the coefficients a_i, viewed as random variables, is small relative to q.

3 Mathematical background

All rings are assumed commutative. An *integral domain* is a unital ring with no nonzero zero-divisors. Among integral domains are *Dedekind domains*, in which all prime ideals are maximal ideals; these include all principal ideal domains (*PID*s), in which every ideal is generated by a single element. Another characterisation of PIDs is that any Dedekind domain that is also a unique factorisation domain is a PID. A special kind of PID is a *Euclidean domain*. Recall that a Euclidean domain is an integral domain A which possesses a function called a *Euclidean function* (or *Euclidean norm*) $d\colon A \setminus \{0\} \to \mathbb{R}_{\geq 0}$ which satisfies

- for all $a, b \in A \setminus \{0\}$, $d(ab) \geq d(a)$;
- for all $a, b \in A$ with $b \neq 0$ there exist $q, r \in A$ such that

$$a = qb + r,$$

  where either $r = 0$ or $d(r) < d(b)$.

For example, $\mathbb{Z}$ equipped with the absolute value is a Euclidean domain, as is the ring $F[X]$ of polynomials over a field equipped with the degree function.

In [11], Kouzmenko discusses how choosing the base ring A to be a Euclidean domain allows one to apply the extended Euclidean algorithm and the Chinese Remainder Theorem in $A[X]/(X^N - 1)$ to invert polynomials modulo a prime in the key creation process of NTRU, but then states that the necessary condition $|p\phi g + fm|_\infty < d(q)$ "does not generally work for all Euclidean rings, since the division algorithm does not produce unique remainder" [11]. In what follows, we clarify the role of having a unique remainder in the Euclidean function (as opposed to in the given implementation of the Euclidean algorithm) and present less restrictive constraints on A allowing its use in an NTRU-like system.

4 NTRU domains

Let A be a ring and set $R = A[X]/(X^N - 1)$. For $b \in A$, let (b) be the ideal generated by b and define $R_b = (A/(b))[X]/(X^N - 1)$. For any $h \in R$, write $[h]_b$ for its image in R_b. Let $p, q \in A$ and $f, g, m, \phi \in R$. To follow an NTRU-like algorithm over A, one must be able to

(1) invert $[f]_q \in R_q$ and $[f]_p \in R_p$;
(2) compute the public key as $[h]_q = [f]_q^{-1}[g]_q \in R_q$;
(3) compute the ciphertext as $[c]_q = [p\phi]_q[h]_q + [m]_q \in R_q$;
(4) compute $[a]_q = [c]_q[f]_q \in R_q$;
(5) identify a representative $a' \in R$ of $[a]_q$;

(6) compute $[m']_p = [a']_p [f]_p^{-1} \in R_p$; and
(7) identify a representative $m' \in R$ of $[m']_p$.

When $a' \neq p\phi g + fm$ (and thus $m' = m$), we say we have a *decryption failure*. Such failures can be exploited to attack NTRU, so it is important to characterize and minimize the probability of decryption failure.

We see that the two obstructions to extending this algorithm to a general ring A in which multiplication and division are efficiently implementable are: inverting elements in Step 1; and identifying representatives in R of elements of R_q and R_p in Steps 5 and 7, respectively. We discuss these issues in Sect. 4.1 and 4.2, below, in the spirit of [11].

4.1 Inverting unit elements in R_q

Theorem 4.1 *Let A be a Dedekind domain and $q \in A$ a prime element such that operations in $A/(q)$ may be efficiently computed. Then there exists an efficient algorithm to determine if $[f]_q \in R_q$ is invertible and to compute its inverse.*

Proof Since q is prime, the ideal (q) is a prime ideal and hence, since A is a Dedekind domain, (q) is a maximal ideal. It follows that $A/(q)$ is a field, and hence the polynomial ring $(A/(q))[X]$ is a Euclidean domain. An element $F \in (A/q)[X]$ (such as $F = [f]_q$) is invertible in $R_q = (A/(q))[X]/(X^N - 1)$ if and only $\gcd(F, X^N - 1) = 1$, and this can be computed by the Euclidean algorithm. The Euclidean algorithm here is based on the degree function in $(A/(q))[X]$; to implement it efficiently requires efficient division and subtraction in $A/(q)$. The extended Euclidean algorithm (which requires products and sums in $A/(q)$) yields $G, H \in (A/(q))[X]$ such that

$$FG + (X^N - 1)H = 1 \in (A/(q))[X];$$

it then follows that $FG = 1 \in R_q$, or $G = F^{-1}$. □

Following the argument used in [11] for the case of $A = \mathbb{Z}[i]$, we have the following immediate extension.

Corollary 4.2 *Let A be a Dedekind domain and $q = \alpha^k \in A$ where α is a prime element. Suppose that operations in A and $A/(\alpha)$ may be efficiently computed. Then there exists an efficient algorithm to determine if $[f]_q \in R_q$ is invertible and to compute its inverse.*

Proof Given $f \in A[X]$, by Theorem 4.1, we can use the Euclidean algorithm in the Euclidean domain $A/(\alpha)[X]$ to determine if $[f]_\alpha$ is invertible in $A/(\alpha)[X]/(X^N - 1)$. If not, then $[f]_q$ is also not invertible, and we are done. If it is, then we can find $G, H \in A/(\alpha)[X]$ such that

$$[f]_\alpha G + (X^N - 1)H = 1 \in A/(\alpha)[X].$$

Choose preimages g and h in $A[X]$ of G and H, respectively. It follows that

$$fg + (X^N - 1)h = 1 - \alpha L$$

for some $L \in A[X]$. Set $\gamma = \prod_{i=0}^{s-1}(1 + (\alpha L)^{2^i})$, where s is the smallest integer such that $2^s > k$. Multiplying both sides by γ and simplifying yields

$$f(\gamma g) + (X^N - 1)\gamma h = 1 - (\alpha L)^{2^s},$$

where the right hand side is equivalent to 1 modulo q and the middle term is equivalent to 0 modulo $X^N - 1$. Hence, the inverse of $[f]_q$ in R_q is

$$\left[\prod_{i=0}^{s-1}\left(1+(\alpha L)^{2^i}\right)g\right]_q.$$

□

The class of Dedekind rings is quite broad. For example, all cyclotomic rings $\mathbb{Z}[\zeta_n]$ are Dedekind rings; as are the rings of algebraic integers of number fields. These latter include such rings as the ring of p-adic integers $\mathbb{Z}_p$, the coordinate rings of suitable non-singular affine curves, and so-called elliptic Dedekind domains, which arise from elliptic curves.

Let us now consider the general case, where $q \in A$ is neither a prime nor a prime power.

Theorem 4.3 *Suppose that A is a PID which has the property that operations in A and in $A/(\alpha)$ for any $\alpha \in A$ can be efficiently computed. Then for any $q \in A$, there exists an efficient algorithm to determine if $[f]_q \in R_q$ is invertible and to compute its inverse.*

Proof Let $q \in A$. Since A is a PID, it is a unique factorization domain, and in particular there exist primes $p_1, p_2, \ldots, p_\ell \in A$, unique up to units, as well as positive integers $e_1, e_2, \ldots, e_\ell$ such that

$$q = p_1^{e_1} p_2^{e_2} \cdots p_\ell^{e_\ell}.$$

In particular, the ideal generated by q factorises as $(q) = \left(p_1^{e_1}\right)\left(p_2^{e_2}\right)\cdots\left(p_\ell^{e_\ell}\right)$. By the Chinese Remainder Theorem, we can write

$$A/(q) \cong A/\left(p_1^{e_1}\right) \times \cdots \times A/\left(p_\ell^{e_\ell}\right).$$

Given $[f]_q \in R_q = A/(q)[X]/(X^N - 1)$, represented by $f \in A[X]$, we may by Corollary 4.2 find for each i an element $g_i \in A/\left(p_i^{e_i}\right)[X]/\left(X^N - 1\right)$ which is an inverse of f modulo $\left(p_i^{e_i}, X^N - 1\right)$ (or else we may conclude that $[f]_q$ is not invertible). Thus the inverse of $[f]_q$ is an element $[g]_q \in R_q$ such that the following systems of equations

$$g \equiv g_i \mod \left(p_i^{e_i}\right) \quad \forall i = 1, \ldots, \ell$$

holds in R. By applying the Chinese Remainder Theorem for PIDs to each coefficient, one deduces that this system has a unique solution in R_q. Calculating this solution explicitly requires an algorithm to compute inverses of elements in each of the quotient rings $A/\left(q/p_i^{e_i}\right)$ of A. □

The set of PIDs is more restrictive than the set of Dedekind domains; for instance, $\mathbb{Z}[\zeta_n]$ is a PID only for 30 values of n, 28 of which are less than 50 [13]. Nevertheless, for any prime p, the ring of p-adic integers $\mathbb{Z}_p$ is a PID; also any polynomial ring $\mathbb{F}[x]$ for $\mathbb{F}$ a field is a PID, as is $\mathbb{Z}$.

Note that the algorithms of Corollary 4.2 and Theorem 4.3 are standard; the key to the suitability of A for NTRU is the existence of an efficient implementation of modular operations.

4.2 Lifting elements from R_q to R

In order to decrypt a message m, we need to choose in Step 5 a value $a \in R$ whose residue mod q satisfies $[a]_q = [c]_q[f]_q = [p\phi g + mf]_q$. Our decryption will be successful exactly if

$$a = p\phi g + mf \in R. \tag{4.1}$$

The difficulty lies in the requirement of defining a function $A/(q) \to A$, which when applied to each of the coefficients of $[a]_q$ gives a maximal probability of successful decryption. (There is also an additional identification modulo p in Step 7; we concentrate on Step 5 as this is the more difficult case.)

Definition 4.4 A subset $S \subset A$ is called a complete set of residues modulo q if

(1) for each $w \in A/(q)$, there exists $s \in S$ so that $[s]_q = w$;
(2) if $s_1, s_2 \in S$ are such that $[s_1]_q = [s_2]_q$ then $s_1 = s_2$.

If $S \subset A$ satisfies (2) but not (1), we call it a partial set of residues modulo q.

Given a complete set of residues modulo q, one has a well-defined injection $\rho : A/(q) \to A$ given by $\rho(w) = s$ whenever $w = [s]_q$. If S is only a partial set of residues, then the map ρ is only partially defined, that is, will fail on certain inputs $w \in A/(q)$. In either case, given such a pair (ρ, S), and $a \in A$, we will have $\rho([a]_q) = a$ if and only if $a \in S$; otherwise, if ρ is defined on $[a]_q$, we have $\rho([a]_q) - a = kq$ for some nonzero $k \in A$.

It follows that one should choose the sets L_m, L_f, L_g and L_ϕ in such a way as to maximize the probability that the coefficients of the product (4.1) will lie in S. An alternative approach would be to choose the sets L_m, L_f, L_g and L_ϕ in a convenient way, and then hope to choose a complete set S of residues based on the probability distribution of the resulting coefficients of a. Our analysis in Sect. 5 combines these two approaches.

Hence, the suitability of a Dedekind domain A for use as a base ring for NTRU hinges on the existence of a good set S of residues modulo q. In the original NTRU, complete sets of residues modulo q are easy to find: any set of q consecutive integers is a complete set of residues. Consequently, given a sufficiently small set of integers Σ (such as those occuring as coefficients of a), one can hope to easily identify a complete set of residues containing Σ.

Let us now consider some concrete examples of domains A which, like $\mathbb{Z}$ admit natural choices of residues S. In particular, we consider Euclidean domains, which satisfy all the conditions laid out in Sect. 4.1, but have additional structure that is useful for defining complete sets of residues.

4.2.1 Lifting elements in a Euclidean domain, part I

Suppose that A is a Euclidean domain with norm d such that given two nonzero elements of A, the remainder and quotient are *uniquely* determined. Although this condition is the one initially thought in [11] as necessary for condition for implementing an NTRU-like cryptosystem, one should note that it is not true of $\mathbb{Z}$, where $n = aq + r = (a + 1)q + (r - q)$ gives two valid choices (in case $r > 0$). In fact, we have the following theorem (see [14], for example).

Theorem 4.5 *Let A be an integral domain that is not a field and that possesses a Euclidean function d such that for every a and non-zero q in A, there exist unique r and s such that*

$$a = rq + s \quad \text{such that} \quad d(s) < d(r) \quad \text{or} \quad s = 0.$$

Then $A = F[x]$ for some field F.

It is then easy to prove the following lemma, where the Euclidean function is the degree function.

Lemma 4.6 *Let F be a field. For $A = F[x]$ and $q \in F[x]\backslash\{0\}$, the set $S = \{s \in A \mid d(s) < d(q)\} \cup \{0\}$ is a complete set of residues modulo q.*

This property is quite desirable. For instance, the CTRU cryptosystem, built upon $A = \mathbb{F}_2[x]$ enjoys this unicity property and consequently has a very simple criteria for successful decryption [5]; unfortunately it was consequently also vulnerable to attack [11], and the given attack would seem to apply to any base field F.

4.2.2 Lifting elements in a Euclidean domain, part II

Now suppose that A is a Euclidean domain which is not a polynomial ring over a field. Then, A possesses a Euclidean norm for which the quotient and remainder are not unique in all cases, so the set $S' = \{s \in A \mid d(s) < d(q)\} \cup \{0\}$, while exhausting all residue classes of $A/(q)$, may contain more than one representative of any particular class in $A/(q)$.

Generally speaking, one might choose a threshold t, depending on q, so that the subset $S = \{s \in A \mid d(s) < t\}$ is a partial set of residues modulo q; this is the approach used in [11], for example.

A more efficient way of choosing a complete set of residues $S \subset S'$ is by making a choice of implementation of the Euclidean algorithm. This is the case for $A = \mathbb{Z}$, for example, where one chooses S to be the subset of S' consisting of all smallest representatives with respect to the Euclidean norm, or more precisely,

$$S = [-q/2, q/2) \cap \mathbb{Z}.$$

This property inspires us to consider $\mathbb{Z}$ as a one dimensional lattice such that the set of residues is the set of lattice points which are closer to 0 than to any point kq on the lattice $(q) = q\mathbb{Z}$. In other words, the Euclidean algorithm is a byproduct of a solution to the closest vector problem (CVP) in the lattice $q\mathbb{Z}$. Generalizing this idea leads us to consider NTRU over Euclidean domains which are themselves lattices in higher dimensions.

One family of lattices which are Dedekind domains are the rings of integers of cyclotomic fields $\mathbb{Q}(\zeta_n)$. They have dimension $\phi(n)$, where ϕ is Euler's totient function. Among these, exactly the rings $\mathbb{Z}[\zeta_n]$ with

$$n \in \{1, 3, 4, 5, 7, 8, 9, 11, 12, 13, 14, 15, 16, 20, 21, 24\}$$

are Euclidean with respect to the restriction of the norm function on $\mathbb{C}$ [12]. Note that $n = 1$ corresponds to the usual ring of integers $\mathbb{Z}$, whereas $\mathbb{Z}[\zeta_4] = \mathbb{Z}[i]$ is the ring of Gaussian integers studied by Kouzmenko in [11].

In this section, we show how the ring of Eisenstein integers $\mathbb{Z}[\zeta_3]$ can be used as an alternative base ring for NTRU; we call the result ETRU.

5 ETRU: NTRU over the Eisenstein integers

We begin by giving an efficient Euclidean algorithm for $\mathbb{Z}[\zeta_3]$, based on its lattice structure, from which follow the efficiency of modular operations in this ring (by the extended Euclidean algorithm).

Recall that $\zeta_3 = \frac{1}{2}(-1 + i\sqrt{3})$ and so $\mathbb{Z}[\zeta_3] \subset \mathbb{C}$. Let us use Greek letters to represent elements of $\mathbb{Z}[\zeta_3]$. The ring $\mathbb{Z}[\zeta_3]$ is a hexagonal lattice in $\mathbb{C} \cong \mathbb{R}^2$ generated by $\{1, \zeta_3\}$. We write $\mathbb{Z}[\zeta_3] = \mathcal{L}(1, \zeta_3)$ when we wish to refer to this basis. Every element $\alpha \in \mathbb{Z}[\zeta_3]$ can be uniquely written as $a + b\zeta_3$ for some $a, b \in \mathbb{Z}$. The (algebraic) norm d of an element $\alpha = a + b\zeta_3 \in \mathbb{Z}[\zeta_3]$ is given by

$$d(a + b\zeta_3) = a^2 + b^2 - ab;$$

this is just the square of the usual (analytic) norm $|\alpha|$ of α viewed as a complex number.

For a nonzero element $\beta \in \mathbb{Z}[\zeta_3]$ the ideal generated by β is

$$\begin{aligned}(\beta) &= \{\gamma\beta \mid \gamma \in \mathbb{Z}[\zeta_3]\} \\ &= \{(a + b\zeta_3)\beta \mid a, b \in \mathbb{Z}\} \\ &= \{a\beta + b\beta\zeta_3 \mid a, b \in \mathbb{Z}\} \\ &= \mathcal{L}(\beta, \beta\zeta_3).\end{aligned}$$

Thus (β) is itself a lattice generated by the scaling (in $\mathbb{Z}[\zeta_3]$) by β of the lattice $\mathcal{L}(1, \zeta_3)$. In particular, it is also a hexagonal lattice, with minimum (analytic) distance $\sqrt{d(\beta)}$ between any two points.

Lemma 5.1 *Given $\alpha \in \mathbb{Z}[\zeta_3]$, if $\gamma\beta$ is the closest lattice point in (β) to α and if $\rho = \alpha - \gamma\beta$, then*

$$\alpha = \gamma\beta + \rho$$

and $d(\rho) < d(\beta)$.

In other words, any closest vector algorithm for the lattice (β) gives an implementation of the Euclidean algorithm for $\mathbb{Z}[\zeta_3]$, due to the relation between the algebraic and analytic norms on $\mathbb{Z}[\zeta_3]$.

Proof By geometric arguments, any point α is at an analytic distance of at most $\sqrt{d(\beta)/3}$ from the closest lattice point in $\mathcal{L}(\beta, \beta\zeta_3)$. Hence $\sqrt{d(\rho)} \leq \sqrt{d(\beta)/3}$ which implies $d(\rho) \leq d(\beta)/3 < d(\beta)$. □

One efficient algorithm to find the closest vector in a lattice is using the Viterbi algorithm on its trellis diagram [3]. To construct the trellis, we choose the rectangular sublattice $\mathcal{L} = \mathcal{L}(\beta, (2\zeta_3 + 1)\beta) = \mathcal{L}(\beta, i\sqrt{3}\beta)$ and note that $\mathcal{L}(\beta, \beta\zeta_3)$ is a union of $\mathcal{L}$ and one of its cosets:

$$\mathcal{L}(\beta, \beta\zeta_3) = \mathcal{L} \cup (\beta\zeta_3 + \mathcal{L}).$$

Since finding the closest point on a rectangular lattice reduces to separately considering each of its coordinates relative to the orthogonal axes of the lattice, this decomposition yields an efficient closest vector algorithm, as described in Theorem 5.2, below.

Given $\beta = c + id \in \mathbb{Z}[\zeta_3]$, define

$$\langle\beta\rangle = \begin{bmatrix} c & -d \\ d & c \end{bmatrix}.$$

For any $z = x + iy \in \mathbb{C}$, we identify it with the vector $\vec{z} = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2$. Thus $\overrightarrow{\beta z} = \langle\beta\rangle\vec{z}$. Let $\lfloor a \rceil$ denote the integer closest to $a \in \mathbb{R}$; if a is equidistant take the smaller of the two choices.

Theorem 5.2 (Closest Lattice Vector Algorithm for (β)) *For any $\alpha = x + iy \in \mathbb{Z}[\zeta_3]$, set $\begin{bmatrix} a_1 \\ b_1 \end{bmatrix} = \langle\beta\rangle^{-1}\vec{\alpha}$ and $\begin{bmatrix} a_2 \\ b_2 \end{bmatrix} = \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} - \vec{\zeta_3}$. For $j = 1, 2$, compute*

$$\rho_j = (a_j - \lfloor a_j \rceil) + i\left(b_j - \sqrt{3}\left\lfloor b_j/\sqrt{3} \right\rceil\right).$$

If $d(\rho_1) \leq d(\rho_2)$ *set* $\vec{\rho} = \langle\beta\rangle\vec{\rho}_1$ *and* $\gamma = (\lfloor a_1 \rceil + i\sqrt{3}\lfloor b_1/\sqrt{3}\rceil)$*; otherwise, set* $\vec{\rho} = \langle\beta\rangle\vec{\rho}_2$ *and* $\gamma = (\lfloor a_2 \rceil + i\sqrt{3}\lfloor b_2/\sqrt{3}\rceil) + \zeta_3$*. Then*

$$\alpha = \gamma\beta + \rho$$

and $\gamma\beta$ *is the closest lattice point in* (β) *to* α*.*

Proof That $\alpha = \beta\gamma + \rho$ in either case follows by expanding the right hand side. Multiplication by $\langle\beta\rangle$ in $\mathbb{R}^2$ is multiplication by β in $\mathbb{C}$; hence up to scaling by $d(\beta)$ it is an orthogonal transformation and thus preserves relative lengths. Hence we may work in coordinates relative to $\{\beta, i\beta\}$. There, $\mathcal{L}(\beta, \beta\zeta_3)$ is the union of the rectangular lattice $\mathcal{L}_R$ spanned by $\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\\sqrt{3}\end{bmatrix}\right\}$ and its coset $\vec{\zeta_3} + \mathcal{L}_R$. The closest point on a rectangular unit lattice is given by rounding off each of the coordinates to the nearest integer; it follows that ρ_1 determines the vector to the closest lattice point on $\mathcal{L}_R$ to $\begin{bmatrix}a_1\\b_1\end{bmatrix}$. Similarly, ρ_2 determines the vector to the closest lattice point on $\mathcal{L}_R$ to $\begin{bmatrix}a_2\\b_2\end{bmatrix}$, or equivalently, the closest lattice point of $\vec{\zeta_3} + \mathcal{L}_R$ to $\begin{bmatrix}a_1\\b_1\end{bmatrix}$. □

Execution of the algorithm of Theorem 5.2 implies a constant number of operations for any β (approximately 10 multiplications and 4 squares), from which we deduce the following corollary.

Corollary 5.3 *The ring* $\mathbb{Z}[\zeta_3]$ *satisfies the hypotheses of Theorems* 4.1 *and* 4.3*, and thus is suitable for use as an NTRU base ring.*

The proof of Lemma 5.1 implies that the Euclidean algorithm of Theorem 5.2 will return a value $\rho \in \mathbb{Z}[\zeta_3]$ in the closure of the Voronoi cell of the origin S. Note that S is a hexagon inscribed between the circles of radius $\frac{1}{2}\sqrt{d(\beta)}$ and $\sqrt{d(\beta)/3}$, that is:

$$\{\rho \in \mathbb{Z}[\zeta_3] \mid d(\rho) \leq d(\beta)/4\} \subset S \subset \{\rho \in \mathbb{Z}[\zeta_3] \mid d(\rho) \leq d(\beta)/3\}. \tag{5.1}$$

The Voronoi cell defines our set of residues; the (topological) interior of S is a partial set of residues in the sense of Definition 4.4, whereas the boundary may contain more than one representative of a given class. However, by our choice of implementation, the algorithm of Theorem 5.2 will return a unique answer.

6 Decryption failure security in ETRU

The level of decryption failure security is measured by the probability that all the (non-modular) coefficients of the polynomial a will lie in the set S of residues (and thus lift to the correct values). In this section, we prove the following proposition.

Proposition 6.1 *Under standard simplifying assumptions, and holding all parameter sizes constant, the decryption failure probability of ETRU is lower than that of NTRU and comparable to that of GNTRU.*

Proof Kouzmenko [11, Chap. 4.4, 6.4] demonstrated the validity of a simple probabilistic model for the coefficients of a. Using this model, we estimate the probability distribution of the coefficients of a, under certain reasonable hypotheses.

We begin with the simplifying assumption (as in [8]) that L_f, L_g and L_ϕ are sets which consist of polynomials with d nonzero coefficients, such that these coefficients are equally distributed among the unit values. In the case of ETRU, for example, we assume they are equally distributed among the six values $\{\pm 1, \pm\zeta_3, \pm\zeta_3^2\}$. For GNTRU and ETRU we choose to model the distribution of these coefficients as complex random variables with mean 0 and variance d/N. For effective comparison with NTRU, we use NTRU at size $2N$ (with coefficients modelled as real random variables of mean 0 and variance d/N). For the purposes of analysis we also model the messages as polynomials whose coefficients are Gaussian random variables with mean 0 and variance ν^2 (real or complex, as appropriate). In what follows, we treat the coefficients of all these polynomials as uncorrelated random variables.

From the decryption algorithm we know that the ith coefficient of a is given by:

$$a_i = p \sum_{k+\ell\equiv i} \phi_k g_\ell + \sum_{k+\ell\equiv i} f_k m_\ell, \quad 0 \le i \le N-1. \tag{6.1}$$

By our assumptions above, it follows that the expected value of this sum is 0; what is of interest is the variance. Recall that the variance of a complex random variable x with mean 0 is given by $E(x\overline{x})$. Thus for ETRU and GNTRU we have

$$\begin{aligned}
\sigma^2 &= E\,(a_i\overline{a_i}) \\
&= E\left[\left(p\sum_{k+\ell\equiv i}\phi_k g_\ell + \sum_{k+\ell\equiv i} f_k m_\ell\right)\left(\overline{p\sum_{k'+\ell'\equiv i}\phi_{k'} g_{\ell'} + \sum_{k'+\ell'\equiv i} f_{k'} m_{\ell'}}\right)\right] \\
&= E\left(|p|^2\sum \phi_k g_\ell \overline{\phi_{k'} g_{\ell'}} + p\sum \phi_k g_\ell \overline{f_{k'} m_{\ell'}} + \overline{p}\sum f_k m_\ell \overline{\phi_{k'} g_{\ell'}} + \sum f_k m_\ell \overline{f_{k'} m_{\ell'}}\right) \\
&= E\left(|p|^2 \sum_{k+\ell\equiv i} |\phi_k|^2\,|g_\ell|^2 + \sum_{k+\ell\equiv i} |f_k|^2\,|m_\ell|^2\right) \quad \text{(since uncorrelated)} \\
&= |p|^2 N\left(\frac{d}{N}\right)\left(\frac{d}{N}\right) + N\left(\frac{d}{N}\right)\nu^2 \\
&= |p|^2 d^2/N + d\nu^2.
\end{aligned}$$

For the case of NTRU, the variance on the individual coefficients is the same but there are $2N$ of them; hence one obtains instead $\sigma^2 = 2(|p|^2 d^2/N + d\nu^2)$.

Let us write β in place of q, for all three systems. The probability of decryption success is given by the probability that each coefficient of a lies in the Voronoi cell of the origin in the lattice generated by β. In the case of NTRU, this is the interval $(-\beta/2, \beta/2)$; for GNTRU this is the square of side length $|\beta|$; for ETRU this is the hexagon with an inscribed circle of radius $|\beta|/2$. Thus one measure is to consider the ratio $|\beta|/\sigma$: the larger this ratio, the more likely it is that the coefficients of a lie in a circle of radius $|\beta|/2$, and hence inside the desired set (in all three cases). By this measure, the decryption failure security of ETRU (and of GNTRU) is slightly greater than that of NTRU. On the other hand, the decryption failure security of ETRU and GNTRU are predicted to be similar. □

From this proposition, we deduce that the decryption failure security of NTRU, ETRU and GNTRU are quite comparable. However, in the most recent publications of security parameters for NTRU, such as [6,10], the value of q (here, β) is chosen large enough so that *no* decryption failure can occur.

From that point of view, and given Proposition 6.1, the two most significant measures of the strength of NTRU are its *combinatorial security* (measured by the size of the key space) and its *lattice security* (measured as a function of the NTRU lattice).

We first note that the hexagonal lattice structure of $\mathbb{Z}[\zeta_3]$ is denser than that of $\mathbb{Z}[i]$ (for GNTRU) or $\mathbb{Z}$ (for NTRU), implying a much larger ring for any given size of parameter. Hence ETRU will have correspondingly higher levels of combinatorial security insofar as the key spaces are larger.

What is more interesting is the issue of lattice security, which we discuss in the next section.

7 Lattice problem for ETRU

Coppersmith and Shamir demonstrated that the security of NTRU (over $\mathbb{Z}$) relies on the hardness of the shortest vector problem in a $2N$-dimensional lattice, as follows. Namely, identify elements of $\mathbb{Z}^N$ with polynomials in $\mathbb{Z}[X]/(X^N - 1)$ (which we denote $\vec{f} \sim f(X)$ if needed, and simply f otherwise); then we consider the $2N$-dimensional lattice in $\mathbb{R}^{2N}$ spanned by the columns of the matrix

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

where H is the circulant matrix corresponding to the public key polynomial h (so $H\vec{f} \sim h(X)f(X)$ if $\vec{f} \sim f(X)$). For a suitable choice of vector u, one has that

$$M \begin{bmatrix} f \\ u \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix}$$

where the pair (f, g) is the private key. Since the vector (f, g) has many zero coefficients, and the nonzero coefficients have norm 1, it is a very short vector in this lattice. More conclusively, Coppersmith and Shamir argue that even if this lattice contains a vector (f', g') shorter than (f, g), then the pair of polynomials (f', g') would function as an alternate key.

To adapt the lattice attack to work with ETRU, we note that $\mathbb{Z}[\zeta_3][X]/(X^N - 1)$ is a $2N$-dimensional real lattice with basis $\{1, \zeta_3, X, \zeta_3 X, X^2, \zeta_3 X^2, \ldots, X^{N-1}, \zeta_3 X^{N-1}\}$. However, in coordinates with respect to this basis, the Euclidean norm of $\zeta_3^2 = -1 - \zeta_3$ is greater than 1, in contradiction to its known analytic and algebraic norm of 1. The basis we should consider instead is the standard (orthogonal) basis

$$\{1, i, X, iX, X^2, iX^2, \ldots, X^{N-1}, iX^{N-1}\} \tag{7.1}$$

of $\mathbb{C}[X]/(X^N - 1)$. With respect to this basis, the Euclidean norm of a polynomial is proportional to the root mean square of the analytic norms of its coefficients, as required. This is the basis used for extending the LLL algorithm to complex vector spaces, as seen in [9, Chap. 3].

Hence to define M we must first apply the change of basis matrix

$$Z := \begin{bmatrix} 1 & -1/2 \\ 0 & \sqrt{3}/2 \end{bmatrix}$$

to each vector $a + b\zeta_3 = \begin{bmatrix} a \\ b \end{bmatrix}$, before constructing the matrix M in the obvious way.

More precisely, for each complex number $z = a + bi$, set $\langle z \rangle := \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$, and for each $n \times n$ matrix A with complex entries, set $\langle A \rangle$ to be the real $2n \times 2n$ matrix obtained by replacing each entry a_{ij} of A with the 2×2 matrix $\langle a_{ij} \rangle$. Thinking of β, and each coefficient of h, as a complex number, we thus can define real $2N \times 2N$ matrices $\langle \beta I_N \rangle$ and $\langle H \rangle$, where H is the circulant matrix as above.

Proposition 7.1 *The ETRU lattice is generated by the columns of the* $4N \times 4N$ *matrix* M*, where*

$$M = \begin{bmatrix} I_{2N} & 0 \\ \langle H \rangle & \langle \beta I_N \rangle \end{bmatrix} \begin{bmatrix} Z & 0 & \cdots & 0 \\ 0 & Z & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & Z \end{bmatrix}.$$

Proof For each vector $\vec{x} \in \mathbb{Z}^{4N}$, write $M\vec{x} = \begin{bmatrix} f \\ g \end{bmatrix}$ where $f, g \in \mathbb{R}^{2N}$. Identifying $\mathbb{C}[X]/(X^N - 1) \simeq \mathbb{R}^{2N}$ via the basis (7.1), we may think of f and g as complex polynomials. Under this identification, it is easy to verify that each column of the product M gives a pair of polynomials whose coefficients lie in $\mathbb{Z}[\zeta_3]$. Since $\begin{bmatrix} f \\ g \end{bmatrix}$ is an integer linear combination of these vectors, the same is true of the corresponding polynomials, and so $f, g \in \mathbb{Z}[\zeta_3][X]/(X^N - 1)$. Furthermore, $\vec{g} = \langle H \rangle \vec{f} + \langle \beta I_N \rangle \vec{u}$ implies that $g = hf + \beta u$ for some polynomial $u \in \mathbb{Z}[\zeta_3][X]/(X^N - 1)$, so the desired relation holds. Finally, the standard norm of a vector in the lattice generated by M equals the square root of the sum of the squares of the analytic norms of each of the coefficients of the corresponding polynomials, so the lengths of these vectors correspond to the size of their corresponding polynomials. □

The distribution of coefficients of f and g in ETRU again ensures that each coefficient is either zero or has the minimal norm of 1, so as with standard NTRU, the private keys (f, g) form short vectors in the lattice, and the same analysis as in [4] applies.

One measure of the security of a lattice $\mathcal{L}$ with respect to solving the shortest vector problem is the lattice constant c; the larger the value of c, the stronger the lattice. The lattice constant is defined by $c = \sqrt{\dim(\mathcal{L})}\ \tau/\sigma$ where τ denotes the length of the actual shortest (nonzero) vector and σ is the Gaussian heuristic estimate of the shortest (nonzero) vector in $\mathcal{L}$, given by

$$\sigma = \sqrt{\dim(\mathcal{L})/2\pi e}\ (\det(\mathcal{L}))^{1/\dim(\mathcal{L})}.$$

Corollary 7.2 *The lattice constant* c_{ETRU} *of the ETRU lattice is, for comparable parameters, 7.5% higher than that of the NTRU or GNTRU lattices.*

Proof Comparing the NTRU, GNTRU and ETRU lattices of given dimension $4N$, with keys chosen at the same norm as in Sect. 6, we see that $c_{ETRU}/c_{NTRU} = (\det(\mathcal{L}_{NTRU})/\det(\mathcal{L}_{ETRU}))^{1/4N}$ and similarly

$$c_{ETRU}/c_{GNTRU} = (\det(\mathcal{L}_{GNTRU}/\det(\mathcal{L}_{ETRU}))^{1/4N}.$$

Now $\det(\mathcal{L}_{NTRU}) = q^{2N}$ and $\det(\mathcal{L}_{GNTRU}) = (|\beta|^2)^N$, whereas

$$\det(\mathcal{L}_{ETRU}) = (|\beta|^2)^N \det(Z)^{2N} = |\beta|^{2N}(\sqrt{3}/2)^{2N}.$$

To hold the decryption failure probability constant, we assume $q = |\beta|$ as in Sect. 6, and deduce that

$$c_{ETRU}/c_{NTRU} = c_{ETRU}/c_{GNTRU} = \sqrt{\frac{2}{\sqrt{3}}} \sim 1.075.$$

□

Finally, we remark that in ETRU there are a number of added difficulties in solving the shortest vector problem. The lattice is not integral, or even rational, due to the non-rationality of the matrix Z. The LLL algorithm applies to any rational approximation to M, but one cannot avoid introducing rounding errors. The problem of choosing an appropriate degree of precision in the approximation to obtain valid results was addressed by Buchmann in [1]. These difficulties imply that the lattice attack on ETRU has a higher cost than on NTRU, or on other integer-lattice based generalizations, like GNTRU.

8 Conclusions and future work

In this paper, inspired by CTRU [5] as well as NTRU over the Gaussian integers [11], we presented possible generalizations of the NTRU cryptosystem to other base rings. We discussed the essential properties of a ring to be used as an NTRU base, concluding in particular that Euclidean domains with lattice structures would be efficient alternatives to explore.

As an example, we showed that $\mathbb{Z}[\zeta_3]$ fulfills these requirements, and moreover, that its lattice structure allows one to easily find a suitable complete set of residues, as is needed in the decryption process. Our theoretical analysis suggests that ETRU would have the same decryption failure security for similar sized parameters. Since $\mathbb{Z}[\zeta_3]$, thought of as a lattice in $\mathbb{C}$, has higher density than the Gaussian integers $\mathbb{Z}[i] \cong \mathbb{Z}^2$ used in [11], ETRU would have a higher level of combinatorial security. Furthermore, the structure of the ETRU lattice is such that its lattice constant c is 7.5% larger than the lattice constants of corresponding NTRU or GNTRU lattices, implying a higher lattice security. Finally, the non-rational lattice structure of $\mathbb{Z}[\zeta_3]$ implies that the standard lattice attacks which apply to NTRU and GNTRU (and upon whose hardness the security of NTRU lies) should be less efficient when attacking ETRU.

Further work on these lines would include presenting an implementation of ETRU to complement the theoretical results presented in this paper. In particular, the closest vector algorithm of Theorem 5.2, as written, is approximately ten times more expensive than similar operations in $\mathbb{Z}$. This is nonetheless competitive, as NTRU claims a 15-fold improvement in speed over ECC for similar security levels [10].

It would be of particular interest to analyse the efficiency and success of attacks using the shortest vector problem in the ETRU lattice, as compared to that for original NTRU at twice the dimension. The theoretical arguments here suggest that the $\mathbb{Z}[\zeta_3]$-lattice would be more secure even while providing a higher data rate. Finally, a number of modifications have been proposed to the original NTRU algorithm, including particular padding schemes, which have in turn proven to be problematic; it would be worthwhile to investigate the possibility of more reliable padding schemes over larger rings such as $\mathbb{Z}[\zeta_3]$.

References

1. Buchmann J.: Reducing lattice bases by means of approximations. Algorithmic number theory, Ithaca, NY, 1994. Lecture Notes in Computer Science, vol. 877, pp. 160–168. Springer, Berlin (1994).
2. Coglianese M., Goi B.-M.: MaTRU: a new NTRU-based cryptosystem. Indocrypt 2005. Lecture Notes in Computer Science, vol. 3797, pp 232–243. (2005).

3. Conway J.H., Sloane N.J.A.: Sphere packings, lattices and groups. In: Grundlehren der Mathematischen Wissenschaften, 3rd edn., vol. 290. Springer-Verlag, New York (1999).
4. Coppersmith D., Shamir A.: Lattice attacks on NTRU. Advances in cryptology—EUROCRYPT 1997. Lecture Notes in Computer Science, vol. 1233, pp. 52–61. Springer, Berlin (1997).
5. Gaborit P., Ohler J., Sole P.: CTRU, A polynomial analogue of NTRU. NTRU Technical Report #Inria RR-4621 (2006).
6. Hirschhorn P., Hoffstein J., Howgrave-Graham N., Whyte W.: Choosing NTRUEncrypt parameters in light of combined lattice reduction and MITM approaches. In: Proceedings of the 7th international conference on applied cryptography and network security, Paris-Rocquencourt, France. Lecture Notes In Computer Science, vol. 5536, pp. 437–455. (2009).
7. Hoffstein J., Pipher J., Silverman J.H.: NTRU, a ring-based public-key cryptosystem. Algorithmic number theory, Portland, OR, 1998. Lecture Notes in Computer Science, vol. 1423, pp. 267–288. Springer, Berlin (1996).
8. Hoffstein J., Pipher J., Silverman J.: An introduction to mathematical cryptography. In: Undergraduate Texts in Mathematics. Springer, New York (2008).
9. Howgrave-Graham N.: Computational mathematics inspired by RSA. PhD thesis, University of Bath (1998).
10. Howgrave-Graham N., Silverman J.H., Whyte W.: Choosing parameter sets for NTRUEncrypt with NAEP and SVES-3. Topics in cryptology—CT-RSA 2005. Lecture Notes in Computer Science, vol. 3376, pp. 118–135. Springer, Berlin (2005).
11. Kouzmenko R.: Generalizations of the NTRU cryptosystem. Diploma Project, École Polytechnique Fédérale de Lausanne, (2005–2006).
12. Lemmermeyer F.: The Euclidean algorithm in algebraic number fields. Exposition. Math. **13**, 385–416, (1995).
13. Masley J.M., Montgomery H.L.: Cyclotomic fields with unique factorization. J. Reine Angew. Math. 248–256 (1976).
14. Rhai T.-S.: A characterization of polynomial domains over a field. Am. Math. Mon. **69**, 984–986 (1962).