Partial Key Exposure Attack on Short Secret Exponent CRT-RSA

Alexander May[1], Julian Nowakowski[1], and Santanu Sarkar[2]

[1] Horst Görtz Institute for IT Security, Ruhr-University Bochum, Germany
{alex.may,julian.nowakowski}@rub.de
[2] Department of Mathematics, Indian Institute of Technology Madras, India
sarkar.santanu.bir@gmail.com

Abstract. Let (N, e) be an RSA public key, where $N = pq$ is the product of equal bitsize primes p, q. Let d_p, d_q be the corresponding secret CRT-RSA exponents.
Using a Coppersmith-type attack, Takayasu, Lu and Peng (TLP) recently showed that one obtains the factorization of N in polynomial time, provided that $d_p, d_q \leq N^{0.122}$. Building on the TLP attack, we show the first *Partial Key Exposure* attack on short secret exponent CRT-RSA. Namely, let $N^{0.122} \leq d_p, d_q \leq N^{0.5}$. Then we show that a constant known fraction of the least significant bits (LSBs) of both d_p, d_q suffices to factor N in polynomial time.
Naturally, the larger d_p, d_q, the more LSBs are required. E.g. if d_p, d_q are of size $N^{0.13}$, then we have to know roughly a $\frac{1}{5}$-fraction of their LSBs, whereas for d_p, d_q of size $N^{0.2}$ we require already knowledge of a $\frac{2}{3}$-LSB fraction. Eventually, if d_p, d_q are of full size $N^{0.5}$, we have to know all of their bits. Notice that as a side-product of our result we obtain a heuristic deterministic polynomial time factorization algorithm on input (N, e, d_p, d_q).

Keywords: CRT-RSA, Coppersmith's method, Partial Key Exposure

1 Introduction

The RSA cryptosystem has the remarkable property that it admits polynomial time attacks for small secrets. Since Wiener's attack [29] for secret exponents $d \leq N^{\frac{1}{4}}$ and Coppersmith's seminal work [6] on factoring $N = pq$ given half of the bits of p, there has been a long line of research on RSA cryptanalysis.

Using Coppersmith's method, Wiener's bound has been improved by Boneh and Durfee [5] to $d \leq N^{0.284}$, respectively $N^{0.292}$, which despite some efforts [16, 26] remains the best known small secret RSA exponent bound. Coron and May [22, 8] proved that on input (N, e, d) the factorization of N can be found in polynomial time.

Afterwards, Ernst, Jochemsz, May, and de Weger [9] showed that both latter results can be linked by a Partial Key Exposure attack. Namely in the range $N^{0.284} \leq d \leq N$, there exists an RSA Partial Key Exposure attack on the most

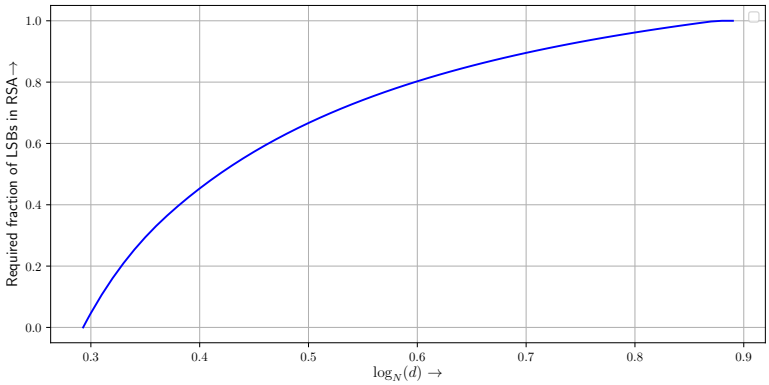

Fig. 1. Required fraction of LSBs for the best known Partial Key Exposure attacks on RSA.

significant bits (MSBs) of d. More precisely, for all d's in this range there is a constant fraction of MSBs whose knowledge allows to factor N in polynomial time. As one would expect, if d is slightly larger than $N^{0.284}$ then one needs only a small MSB bit fraction, whereas for d tending to N (or more precisely $\phi(N)$) one needs all of d's bits.

Later this Partial Key Exposure attack was improved by Takayasu and Kunihiro [24] to cover the range $N^{0.292} \leq d \leq N$ of the superior Boneh-Durfee bound. Notice that for Partial Key Exposure attacks a smaller range is indeed an improvement. Whereas in the range $d \in [N^{0.284}, N^{0.292}]$ the attack of [9] requires some known bits, the attack of Takayasu and Kunihiro [24] succeeds in this range without any bit-knowledge. The fact that the superior Boneh-Durfee bound $d \leq N^{0.292}$ extrapolates smoothly to full size $d \leq N$ gives us some indication that [24] might be optimal.

Takayasu and Kunihiro [24] also presented an LSB attack, based on a result by Aono [1], that works in the range $N^{0.292} \leq d \leq N^{0.89}$, see Figure 1. Somewhat surprisingly, it is open whether there exists an LSB-type Partial Key Exposure attack up to full size d.

In practice, RSA Partial Key Exposure attacks led to a wide range of devastating attacks [11, 2, 23] on real-world RSA implementations that leaked private key bits.

CRT-RSA. As opposed to small secret d, the case of small CRT exponents seems to be notoriously harder to analyze. The existence of such attacks was initially raised as an open problem in Wiener [29]. The first result was achieved in [20] only for primes p, q of imbalanced bitsize, and later improved in [3]. The first bound for the standard RSA setting with balanced primes was given by Jochemsz and May [15], who showed a Coppersmith-type polynomial time

attack for $d_p, d_q \leq N^{0.073}$. This was recently improved by Takayasu, Lu and Peng [27] to $N^{0.091}$ and shortly after [28] to a remarkably large bound $N^{0.122}$. We refer to the latter bound as the TLP attack.

However, several natural questions remain unanswered. First, the optimality of the TLP attack is unclear, especially since TLP is a highly involved application of Coppersmith's method to a system of three polynomials. Second, it remained open whether small CRT exponents admit Partial Key Exposure attacks at all. Partial Key Exposure attacks on CRT exponents where so far only known for the special setting of small public exponents e, see [4, 18, 25]. And third, even if small CRT exponent Partial Key Exposure attacks exist, do they interpolate to the natural bound $d_p, d_q \leq N^{0.5}$? For this bound, i.e. known CRT-exponents, Maitra and Sarkar [19] showed a deterministic Coppersmith-type factorization attack on input (N, e, d_p, d_q).

Our results. As our main result, we give the first Partial Key Exposure attack on CRT exponents in the full range $N^{0.122} \leq d_p, d_q \leq N^{0.5}$, see Figure 2 for an illustration. Since we achieve a smooth interpolation from the TLP result $N^{0.122}$ to the natural upper bound $N^{0.5}$, this gives some indication of optimality. Our upper bound provides a *heuristic* deterministic polynomial time factorization algorithm on input (N, e, d_p, d_q), different from the one of Maitra and Sarkar [19]. For our results, we require the typical well-studied Coppersmith heuristic for multivariate polynomials, as e.g. used in [1, 3–5, 9, 12, 15, 16, 19, 28].

On the way to achieving our main result, we make some contributions that might be of independent interest. First, we give a geometric interpretation of the TLP attack in terms of Newton polytopes that helps to gain a deeper structural insight. Second, we show a simplified LSB Partial Key Exposure attack in the range $N^{0.083} \leq d_p, d_q \leq N^{0.5}$, see Figure 2.

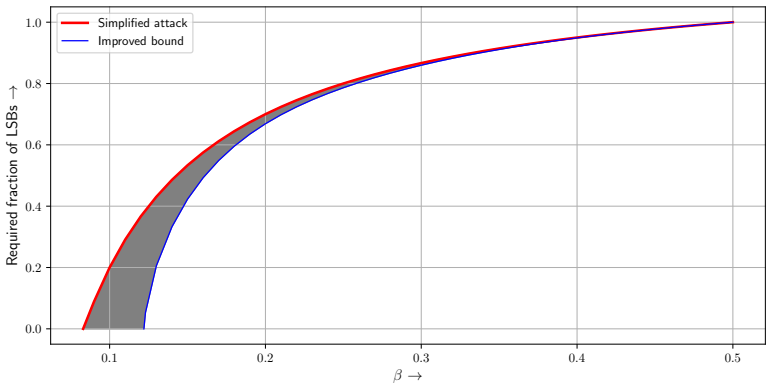

Fig. 2. Comparison between our simplified attack and our main result.

This attack admits an elegant formula as follows. Assume that d_p, d_q are of size N^β and write $d_p = d_p^* 2^k + \widetilde{d_p}$, $d_q = d_q^* 2^k + \widetilde{d_p}$ for some k, known LSBs $\widetilde{d_p}, \widetilde{d_q}$, and unknown MSBs $d_p^*, d_q^* \leq N^\delta$. Then we can find the factorization of N in polynomial time under the usual Coppersmith-type heuristic, provided that $\delta \leq \frac{1}{10} - \frac{1}{5}\beta$.

Notice that our formula already has the desired end point $d_p, d_q \leq N^{\frac{1}{2}}$. For any $\beta \leq \frac{1}{2}$, i.e., for any d_p, d_q up to full size, we obtain a non-negative bound for δ. For $\beta = \delta$, in which case we do not know any LSBs, we achieve $\delta \leq \frac{1}{12} \approx 0.083$.

Eventually, we optimize our attack such that it works in the range $N^{0.122} \leq d_p, d_q \leq N^{1/2}$, i.e., building on top of the TLP bound. This improves on our simplified Partial Key Exposure attack, since it requires no key-knowledge in the range $d_p, d_q \in [N^{0.083}, N^{0.122}]$. Moreover, for any secret exponent size in the range $N^{0.083} \leq d_p, d_q \leq N^{1/2}$ it requires less key-knowledge of d_p, d_q, see Figure 2 for a comparison of the required LSB fraction.

We find it somewhat remarkable that our CRT-RSA LSB attack works for full size d_p, d_q, whereas the best known RSA LSB Partial Key Exposure attack [24] from Figure 1 does not reach full size d.

Since RSA Partial Key Exposure attacks already found many real-world applications [11, 2, 23], we hope that our CRT-RSA counterpart also stimulates further research in this area. We believe that in practice bits of d_p, d_q might be easier to get via side-channel attacks than bits of d, since almost all standard RSA implementations for efficiency reasons actually use CRT exponents.

Our paper is structured as follows. In Section 2, we recall the basics of Coppersmith's method. In Section 3, we revisit the TLP attack, and thoroughly analyze TLP using our new geometric approach. This reformulation then in turn allows us to easily prove our simplified small CRT exponent attack in Section 4. To show our main result for the improved CRT attack in the range $N^{0.122} \leq d_p, d_q \leq N^{0.5}$ in Section 4.1, we again heavily reuse our results from Section 3. We conclude by providing experimental evidence of our standard Coppersmith-type heuristic in Section 5.

2 Coppersmith's Method

Like in many other attacks on RSA, we base our attack on Coppersmith's method for finding small modular roots of multivariate polynomials [7]. For that, we model the problem of factoring an RSA modulus as a problem of finding a small root of multivariate polynomials modulo some large integer M. In particular, we use the RSA key generation equations to derive n polynomials $f_1, \ldots, f_n$ in k variables $x_1, \ldots, x_k$, which share a small root $r = (r_1, \ldots, r_k)$ modulo M. Small means here that we know for $j = 1, \ldots, k$ upper bounds X_j with $|r_j| \leq X_j$. Then, we chosse an $m \in \mathbb{N}$ and define so-called *shift polynomials*

$$p_i := f_1^{i_1} \cdot \ldots \cdot f_n^{i_n} \cdot x_1^{j_1} \cdot \ldots \cdot x_k^{j_k} \cdot M^{m-(i_1+\ldots+i_n)},$$

with appropriately chosen exponents. Notice that by construction the shift polynomials have the root r modulo M^m.

Our goal is to compute integer linear combinations

$$h_j(x_1, \ldots, x_k) := \sum_i \alpha_{j,i} p_i(x_1, \ldots, x_k) \qquad (\alpha_{j,i} \in \mathbb{Z})$$

of the shift polynomials, to obtain k polynomials $h_1, \ldots, h_k$, such that for every $j = 1, \ldots, k$ the coefficient vector of $h_j(X_1 x_1, \ldots, X_k x_k)$ has sufficiently small Euclidean norm. A lemma by Howgrave-Graham (as stated below) then guarantees us that $h_1, \ldots, h_k$ have the root r not just modulo M^m, but also over the integers. If the variety of the ideal $(h_1, \ldots, h_k)$ is zero-dimensional, this allows us to recover their root by using a Groebner basis – which in our case means that we can efficiently factor the RSA modulus.

Lemma 1 (Howgrave-Graham, [14]). *Let $h(x_1, \ldots, x_k) \in \mathbb{Z}[x_1, \ldots, x_k]$ be a polynomial in at most ω monomials. Suppose that $h(r_1, \ldots, r_k) \equiv 0 \bmod M^m$ for some positive integer m. Also let $|r_i| < X_i$ for $1 \leq i \leq k$ and*

$$||h(x_1 X_1, \ldots, x_k X_k)|| < \frac{M^m}{\sqrt{\omega}}.$$

Then $h(r_1, \ldots, r_k) = 0$ holds over the integers.

To find suitable polynomials h_j, we use lattice-based techniques.

Definition 1. *Let $\{\mathbf{b_1}, \ldots, \mathbf{b}_\omega\} \subset \mathbb{Z}^n$ be linearly independent row vectors. The* lattice *$\mathcal{L}$ generated by these vectors is defined by*

$$\mathcal{L} = \left\{ z_1 \mathbf{b_1} + \ldots + z_\omega \mathbf{b}_\omega | z_i \in \mathbb{Z}, \forall i \in \{1, \ldots, \omega\} \right\}.$$

$\{\mathbf{b_1}, \ldots, \mathbf{b}_\omega\}$ is called a basis *of $\mathcal{L}$. The parameter n is called the* dimension *of $\mathcal{L}$, ω is called the rank of $\mathcal{L}$. If $\omega = n$, then we call $\mathcal{L}$ a* full-rank lattice.

We often associate a lattice with a *basis matrix* $\mathbf{B}$. Two lattice bases generate the same lattice if and only if their basis matrices $\mathbf{B}_1$ and $\mathbf{B}_2$ satisfy $\mathbf{B}_1 = \mathbf{U}\mathbf{B}_2$ for some unimodular matrix $\mathbf{U}$. As unimodular square matrices have determinant ± 1, one can define the *determinant of a full-rank lattice* $\mathcal{L}$ as

$$\det \mathcal{L} := |\det \mathbf{B}| .$$

Notice that the coefficient vectors of the polynomials $h_j(X_1 x_1, \ldots, X_k x_k)$, as defined above, are elements of a lattice $\mathcal{L}_S$, which is generated by the coefficient vectors of the polynomials $p_i(X_1 x_1, \ldots, X_k x_k)$. Hence, the problem of finding polynomials h_j with short norm boils down to finding short non-zero vectors in $\mathcal{L}_S$. This can be achieved in polynomial time using the well-known LLL algorithm [17].

Lemma 2. *Let $\mathcal{L}$ be an integer lattice of dimension ω. The LLL algorithm applied to $\mathcal{L}$ outputs a reduced basis $\{\mathbf{v_1}, \ldots, \mathbf{v}_\omega\}$ of $\mathcal{L}$ with*

$$||\mathbf{v_1}|| \leq ||\mathbf{v_2}|| \leq \cdots \leq ||\mathbf{v_i}|| \leq 2^{\frac{\omega(\omega-1)}{4(\omega+1-i)}} \det(\mathcal{L})^{\frac{1}{\omega+1-i}}, \text{ for } i = 1, \ldots, \omega,$$

in time polynomial in the dimension ω and the bit size of the entries of $\mathcal{L}$.

For a proof of Lemma 2, we refer to [21, Theorem 4].

As a consequence of Lemma 2, if the condition

$$2^{\frac{\omega(\omega-1)}{4(\omega+1-\ell)}} \det(\mathcal{L}_s)^{\frac{1}{\omega+1-\ell}} < \frac{M^m}{\sqrt{\omega}},$$

holds for all $\ell \leq k$, we can obtain the required k polynomials h_j, which satisfy the condition of Lemma 1, by simply applying LLL to the lattice $\mathcal{L}_S$. Since in our case the values of the determinant and of M grow significantly faster than the other terms (as usual in these types of attacks), we can also use the simplified *enabling condition*

$$\det \mathcal{L}_S < (M^m)^{\dim \mathcal{L}_S}. \tag{1}$$

To keep the calculation of the determinant simple, we require that the basis matrix of $\mathcal{L}_S$ is of a triangular shape. For that, we need to ensure that the shift polynomial p_1 has exactly one monomial and moreover that for every $i > 1$ the set

$$\{\lambda \mid \lambda \text{ is a monomial of } p_i \text{ but not of } p_1, \ldots, p_{i-1}\}$$

contains exactly one element. Calculating the determinant then becomes particularly easy, as we simply have to keep track for every i, which monomial λ_i the polynomial p_i adds to the basis matrix' diagonal. Denoting the coefficient of λ_i by c_i, the determinant then can be calculated as

$$\det \mathcal{L}_S = \prod_i |c_i \cdot \lambda_i(X_1, \ldots, X_k)| \,.$$

For constructing our basis matrix, we will often make use of a powerful tool, the so called *Newton polytope* of a polynomial.

Definition 2. *The* Newton polytope *of a k-variate polynomial $p(x_1, \ldots, x_k)$ is defined as the convex hull of the set*

$$N(p) := \left\{(i_1, \ldots, i_k) \in \mathbb{N}^k \mid x_1^{i_1} \cdot \ldots \cdot x_k^{i_k} \text{ is a monomial of } p\right\}.$$

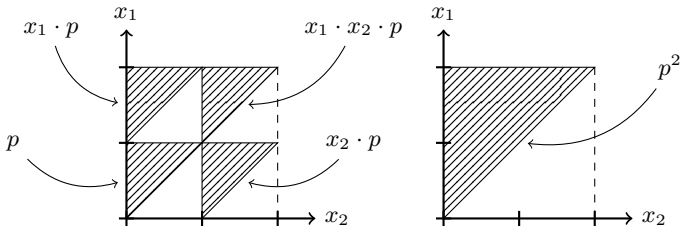

Fig. 3. The Newton polytopes of $p(x_1, x_2) := x_1x_2 + x_1 + 1$ and related polynomials.

Notice that for two polynomials p_1, p_2 the sets $N(p_1)$, $N(p_2)$ as defined above have the useful property that $N(p_1 p_2) = N(p_1) + N(p_2)$, where + denotes the Minkowski sum. Hence, the Newton polytope of some polynomial $x_i^a \cdot p$ (where $a \in \mathbb{N}$) is obtained by moving the Newton polytope of p up a units on the axis corresponding to x_i. Similarly, the Newton polytope of p^a is obtained by scaling the Newton polytope of p by a factor of a. (See Figure 3 for examples.)

It is worth to note that we have no provable guarantee that the LLL gives us polynomials, which generate an ideal with zero-dimensional variety. Thus, our approach relies on the standard Coppersmith-type heuristic assumption.

Assumption 1 *In this work, the lattice based constructions yield polynomials, that generate an ideal with zero-dimensional variety.*

In Section 5 we verify Assumption 1 experimentally.

3 The TLP Attack Revisited

As our attack is strongly based on the Takayasu-Lu-Peng attack (TLP) [28] on CRT-RSA, we describe it in this section in detail. We deviate from the original algebraic TLP formulation, with the hope that our geometric view helps to gain a deeper understanding. We first present a simplified construction and after that optimize it to obtain TLP.

3.1 A Simplified Construction

Let us recall the CRT-RSA key generation equations

$$ed_p = k(p-1) + 1, \tag{2}$$
$$ed_q = \ell(q-1) + 1, \tag{3}$$

where $N = pq$ is an RSA modulus, e is a public exponent, d_p, d_q are the corresponding CRT-exponents and $k, \ell \in \mathbb{N}$. Writing $e = N^\alpha$ and upper bounding $d_p, d_q \leq N^\delta$ for some $\alpha, \delta \in \mathbb{R}$, the values of k and ℓ can be bounded as

$$k = \frac{ed_p - 1}{p-1} < \frac{ed_p}{p-1} = \Theta\left(\frac{ed_p}{N^{1/2}}\right) = \Theta(N^{\alpha+\delta-1/2}),$$
$$\ell = \frac{ed_q - 1}{q-1} < \frac{ed_q}{q-1} = \Theta\left(\frac{ed_q}{N^{1/2}}\right) = \Theta(N^{\alpha+\delta-1/2}),$$

since in the usual RSA setting we have $p, q = \Theta(N^{1/2})$. By that, we find an $X = \Theta(N^{\alpha+\delta-1/2})$, which is an upper bound for both k and ℓ.

We use equation (2) to derive a polynomial

$$f(x_p, y_p) := x_p(y_p - 1) + 1 = x_p y_p - x_p + 1,$$

which has the root (k, p) modulo e. Similarly, we could also use equation (3) to derive another polynomial, which in turn has the root (ℓ, q) modulo e. Takayasu,

Lu and Peng, however, advise to first multiply equation (3) with p and rearrange terms as suggested by Bleichenbacher and May [3]:

$$ped_q = p\ell(q-1) + p = N\ell - p\ell + p = N(\ell-1) + N - p(\ell-1).$$

Then, the equation yields a polynomial

$$g(y_p, z_p) := y_p z_p - N z_p - N,$$

which has the root $(p, \ell-1)$ modulo e.

The multiplication with p has the advantage that we can get rid of the unknown q and by that treat f and g as three-variate polynomials in the variables x_p, y_p, z_p, which have a common root $(k, p, \ell-1)$. Using $\ell-1$ instead of ℓ, gives g a superior Newton polytope, since f and g then share a monomial (see Figure 4).

With f, we now have a polynomial, which relates the unknowns k and p, while g relates ℓ and p. To obtain a third polynomial, that relates k and ℓ, one can use an idea by Galbraith, Heneghan and McKee [10]. First, we rewrite equations (2) and (3) as

$$\begin{aligned} kp &= k - 1 + ed_p, \\ \ell q &= \ell - 1 + ed_q. \end{aligned}$$

Then, multiplying kp with ℓq, we obtain

$$k\ell N = (k-1)(\ell-1) + (k-1)ed_q + ed_p(\ell-1) + e^2 d_p d_q$$

and equivalently

$$(N-1)k(\ell-1) + Nk + (\ell-1) = e\left(d_q(k-1) + d_p(\ell-1) + ed_p d_q\right),$$

from which we can derive a polynomial

$$h(x_p, z_p) := (N-1)x_p z_p + N x_p + z_p$$

with the root $(k, \ell-1)$ modulo e.

Now, we have the following system of polynomial equations

$$\begin{aligned} f(x_p, y_p, z_p) &= x_p y_p - x_p + 1 = 0, \\ g(x_p, y_p, z_p) &= y_p z_p - N z_p - N = 0, \\ h(x_p, y_p, z_p) &= (N-1)x_p z_p + N x_p + z_p = 0, \end{aligned}$$

with the solution $(x_0, y_0, z_0) = (k, p, \ell-1)$ modulo e, which can be upper bounded as

$$\begin{aligned} x_0, z_0 \leq X &= \Theta(N^{\alpha+\delta-1/2}), \\ y_0 \leq Y &= \Theta(N^{1/2}). \end{aligned}$$

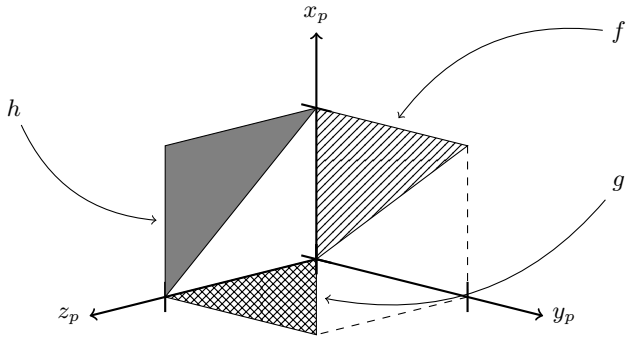

Fig. 4. The Newton polytopes of f, g and h.

If we can efficiently compute (x_0, y_0, z_0), we factor the RSA modulus N.

We want to use Coppersmith's method to compute (x_0, y_0, z_0). For that, we define shift polynomials, which have the root (x_0, y_0, z_0) modulo e^{2m} for some $m \in \mathbb{N}$. The polynomials will form a lattice with triangular lattice basis matrix whose columns correspond to the elements of the set

$$\mathcal{M} := \left\{ x_p^a y_p^b z_p^c \mid x_p^a y_p^b z_p^c \text{ is a monomial of } f^m g^m \right\}.$$

Notice that by Figure 4 we may equivalently define $\mathcal{M}$ as

$$\mathcal{M} = \left\{ x_p^a y_p^b z_p^c \mid 0 \le a \le m, 0 \le c \le m, 0 \le b \le a + c \right\}. \tag{4}$$

We partition $\mathcal{M}$ into four subsets

$$\begin{aligned}
\mathcal{M}_1 &:= \left\{ x_p^a y_p^b z_p^c \in \mathcal{M} \mid a \le c, b \le c - a \right\}, \\
\mathcal{M}_2 &:= \left\{ x_p^a y_p^b z_p^c \in \mathcal{M} \mid a > c, b < a - c \right\}, \\
\mathcal{M}_3 &:= \left\{ x_p^a y_p^b z_p^c \in \mathcal{M} \mid x_p^a y_p^b z_p^c \notin (\mathcal{M}_1 \cup \mathcal{M}_2), a + b + c \equiv 0 \mod 2 \right\}, \\
\mathcal{M}_4 &:= \left\{ x_p^a y_p^b z_p^c \in \mathcal{M} \mid x_p^a y_p^b z_p^c \notin (\mathcal{M}_1 \cup \mathcal{M}_2 \cup \mathcal{M}_3) \right\}.
\end{aligned}$$

These partitions are used to define a collection of functions, which we call the *exponent functions*.

$$E_f(a,b,c) := \begin{cases} 0, & x_p^a y_p^b z_p^c \in \mathcal{M}_1 \\ b, & x_p^a y_p^b z_p^c \in \mathcal{M}_2 \\ (a+b-c)/2, & x_p^a y_p^b z_p^c \in \mathcal{M}_3 \\ (a+b-c+1)/2, & x_p^a y_p^b z_p^c \in \mathcal{M}_4 \end{cases},$$

$$E_g(a,b,c) := \begin{cases} b, & x_p^a y_p^b z_p^c \in \mathcal{M}_1 \\ 0, & x_p^a y_p^b z_p^c \in \mathcal{M}_2 \\ (-a+b+c)/2, & x_p^a y_p^b z_p^c \in \mathcal{M}_3 \\ (-a+b+c-1)/2, & x_p^a y_p^b z_p^c \in \mathcal{M}_4 \end{cases},$$

$$E_h(a,b,c) := \begin{cases} a, & x_p^a y_p^b z_p^c \in \mathcal{M}_1 \\ c, & x_p^a y_p^b z_p^c \in \mathcal{M}_2 \\ (a-b+c)/2, & x_p^a y_p^b z_p^c \in \mathcal{M}_3 \\ (a-b+c-1)/2, & x_p^a y_p^b z_p^c \in \mathcal{M}_4 \end{cases},$$

$$E_x(a,b,c) := \begin{cases} a-b-c, & x_p^a y_p^b z_p^c \in \mathcal{M}_2 \\ 0, & x_p^a y_p^b z_p^c \in \mathcal{M}_1 \cup \mathcal{M}_3 \cup \mathcal{M}_4 \end{cases},$$

$$E_z(a,b,c) := \begin{cases} -a-b+c, & x_p^a y_p^b z_p^c \in \mathcal{M}_1 \\ 0, & x_p^a y_p^b z_p^c \in \mathcal{M}_2 \cup \mathcal{M}_3 \\ 1, & x_p^a y_p^b z_p^c \in \mathcal{M}_4 \end{cases}.$$

One can easily verify that the exponent functions satisfy the following properties.

Lemma 3. *Let $x_p^a y_p^b z_p^c \in \mathcal{M}$. Then the following holds:*

1. $E_f(a,b,c), E_g(a,b,c), E_h(a,b,c), E_x(a,b,c), E_z(a,b,c) \in \mathbb{N}$.
2. $E_f(a,b,c) + E_g(a,b,c) + E_h(a,b,c) \leq 2m$.
3. $E_f(a,b,c) + E_h(a,b,c) + E_x(a,b,c) = a$.
4. $E_f(a,b,c) + E_g(a,b,c) = b$.
5. $E_g(a,b,c) + E_h(a,b,c) + E_z(a,b,c) = c$.

Proof. Simply compare the definitions of $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, \mathcal{M}_4$ with those of the exponent functions. □

For a given monomial $x_p^a y_p^b z_p^c \in \mathcal{M}$ we use the exponent functions to define a shift polynomial as follows:

$$\begin{aligned} p_{[a,b,c]}(x_p, y_p, z_p) := f^{E_f(a,b,c)} \cdot g^{E_g(a,b,c)} \cdot h^{E_h(a,b,c)} \cdot \\ x_p^{E_x(a,b,c)} \cdot z_p^{E_z(a,b,c)} \cdot \\ e^{2m-(E_f(a,b,c)+E_g(a,b,c)+E_h(a,b,c))}. \end{aligned}$$

Notice that the first two statements in Lemma 3 ensure that every exponent in $p_{[a,b,c]}$ has a non-negative value. Further notice that $p_{[a,b,c]}$ has the root $(k, p, \ell - 1)$ modulo e^{2m}.

We equip our shift polynomials with the *lexicographic monomial order on* (z_p, x_p, y_p), which in the following we simply call the (z_p, x_p, y_p)*–order.*

Definition 3 ((z_p, x_p, y_p)-order). *The monomial order*

$$x_p^{a_1} y_p^{b_1} z_p^{c_1} < x_p^{a_2} y_p^{b_2} z_p^{c_2} :\Longleftrightarrow \begin{cases} c_1 < c_2 \\ c_1 = c_2, a_1 < a_2 \\ c_1 = c_2, a_1 = a_2, b_1 < b_2 \end{cases}$$

is called the (z_p, x_p, y_p)-order.

The shift polynomials have the following nice properties.

Lemma 4. *Let $x_p^a y_p^b z_p^c \in \mathcal{M}$. Then the following holds:*

1. *The leading monomial of* $p_{[a,b,c]}$ *in the* (z_p, x_p, y_p)*-order is* $x_p^a y_p^b z_p^c$.
2. *The monomials of* $p_{[a,b,c]}$ *form a subset of* $\mathcal{M}$.

Proof. Every shift polynomial is of the form $p_{[a,b,c]} = f^{i_1} g^{i_2} h^{i_3} x_p^{j_1} z_p^{j_2} e^{j_3}$, where the exponents are defined by our exponent functions. From Figure 4, we conclude that the leading monomials of f^{i_1}, g^{i_2} and h^{i_3} are $x_p^{i_1} y_p^{i_1}$, $y_p^{i_2} z_p^{i_2}$ and $x_p^{i_3} z_p^{i_3}$ respectively. Thus, $p_{[a,b,c]}$ has leading monomial

$$x_p^{i_1+i_3+j_1} y_p^{i_1+i_2} z_p^{i_2+i_3+j_2}.$$

Since from Lemma 3 it follows that the exponent functions are defined in such a way that $a = i_1 + i_3 + j_1$, $b = i_1 + i_2$ and $c = i_2 + i_3 + j_2$ always holds, this proves the first statement in the lemma.

To prove the second statement, we conclude from Figure 4 that the set of the monomials $p_{[a,b,c]}$ is a subset of

$$\mathcal{M}' := \left\{ x_p^{a'} y_p^{b'} z_p^{c'} \mid 0 \leq a' \leq i_1 + i_3 + j_1, 0 \leq c' \leq i_2 + i_3 + j_2, 0 \leq b' \leq a' + c' \right\}.$$

Thus, it suffices to show that $\mathcal{M}' \subseteq \mathcal{M}$.

From the above, we conclude

$$x_p^{i_1+i_3+j_1} y_p^{i_1+i_2} z_p^{i_2+i_3+j_2} = x_p^a y_p^b z_p^c \in \mathcal{M}.$$

Hence, from (4) it follows that $i_1 + i_3 + j_1 \leq m$ and $i_2 + i_3 + j_2 \leq m$. Comparing the definition of $\mathcal{M}'$ with (4), the statement $\mathcal{M}' \subseteq \mathcal{M}$ easily follows. □

Using Lemma 4 we now prove the following important proposition.

Proposition 1. *Order the monomials in* $\mathcal{M}$ *according to the* (z_p, x_p, y_p)*-order. Define a lattice basis matrix* $\mathbf{B}$*, in which the* i*-th column corresponds to the* i*-th smallest monomial* $x_p^a y_p^b z_p^b \in \mathcal{M}$ *and the* i*-th row corresponds to the coefficient vector of the polynomial* $p_{[a,b,c]}(Xx_p, Yy_p, Xz_p)$*. Then* $\mathbf{B}$ *is triangular.*

Proof. If $p_{[a,b,c]}$ has a monomial $x_p^{a'} y_p^{b'} z_p^{c'} \neq x_p^a y_p^b z_p^c$, then with Lemma 4 it follows that $x_p^{a'} y_p^{b'} z_p^{c'} < x_p^a y_p^b z_p^c$ and furthermore $x_p^{a'} y_p^{b'} z_p^{c'} \in \mathcal{M}$. Therefore, when adding $p_{[a,b,c]}$ to $\mathbf{B}$, $x_p^{a'} y_p^{b'} z_p^{c'}$ already is included, as it is the leading monomial of some polynomial $p_{[a',b',c']}$, which, by construction, is added before $p_{[a,b,c]}$ to $\mathbf{B}$. Conversely, no polynomial $p_{[a',b',c']}$, which is added before $p_{[a,b,c]}$ to $\mathbf{B}$, has the monomial $x_p^a y_p^b z_p^c$, since all its monomials are strictly smaller than $x_p^a y_p^b z_p^c$. Hence, $p_{[a,b,c]}$ has with $x_p^a y_p^b z_p^c$ exactly one monomial, which is not added priorly to the basis. □

In Figure 5 we give an example of the lattice construction as described in Proposition 1 for the case $m = 2$. The table on the left shows the polynomials, that are included in the lattice. The table on the right shows the corresponding leading monomials. The cell colours indicate, in which set $\mathcal{M}_i$ the leading monomials lie. For the sake of a simpler notation, we omit the powers of e that are multiplied to the shift polynomials.

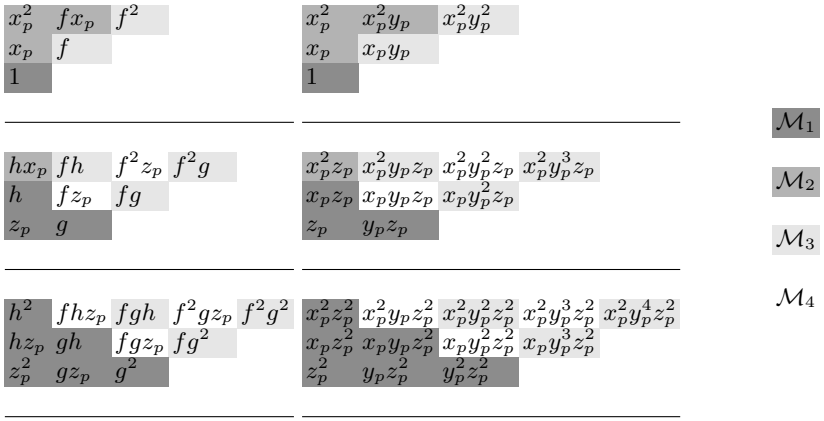

Fig. 5. The lattice construction as described in Proposition 1 for $m = 2$.

The entry in the a-th row of the b-th column in the c-th block corresponds to the shift polynomial $p_{[m-a,b,c]}$. (We chose to use $m-a$ instead of a, as the shape of the tables then matches the shape of the Newton polytope of $f^m g^m$.) Notice that the monomials in $\mathcal{M}_1$ and $\mathcal{M}_2$ are added to the lattice by polynomials, which contain only powers of g, h and z_p or f, h and x_p respectively. The monomials in $\mathcal{M}_3$ and $\mathcal{M}_4$ are added by multiplying powers of f to the polynomials, that lie on the right border of the lower triangles corresponding to $\mathcal{M}_1$.

Remark 1. We would like to explain the optimization process, that led us to the definitions of the exponent functions. To keep the lattice's determinant as small as possible, the sum

$$E_f(a,b,c) + E_g(a,b,c) + E_h(a,b,c)$$

should be maximized for every shift polynomial $p_{[a,b,c]}$. (The larger the sum, the smaller the power of e in the shift polynomial and by that the value of the determinant.) If one wants to use shift polynomials, which satisfy the useful properties of Lemma 4, then with Figure 4 it is not hard to see that the optimal values for the exponent functions are obtained by maximizing the sum under the constraints

$$\begin{aligned} E_f(a,b,c) + E_h(a,b,c) &\leq a, \\ E_f(a,b,c) + E_g(a,b,c) &\leq b, \\ E_g(a,b,c) + E_h(a,b,c) &\leq c. \end{aligned}$$

This suggests that the problem of selecting optimal exponent functions can be modelled as an integer programming problem. We solved the integer programming problem for efficiently solvable instances of a, b and c, looked for patterns in its solutions and then based the definitions of the exponent functions on those.

For all instances of a, b and c, that we checked, our definitions perfectly match the optimal solution of the corresponding integer programming problem. This gives some evidence for the optimality of our definitions.

Unfortunately, our lattice construction so far does not result in a successful attack, as for any value of m it does not satisfy the enabling condition (1). In fact, no shift polynomial in our lattice is *helpful*, since no polynomial adds a factor smaller than e^{2m} to the lattice's determinant. However, as we will see below, by only slightly enhancing the construction with some clever tricks as suggested by Takayasu, Lu and Peng in [28], we immediately obtain their lattice, which then yields the attack that works whenever $\delta < 0.122$.

3.2 Improving the Construction via Unravelled Linearization

Instead of using three-variate shift polynomials in the variables x_p, y_p, z_p, we now want to use six-variate polynomials in the variables $x_p, x_q, y_p, y_q, z_p, z_q$, which have the root $r := (k, k-1, p, q, \ell-1, \ell)$ modulo e^{2m}. With these new variables, we can apply *unravelled linearization* as introduced by Hermann and May [12, 13] to our polynomials. That is, we can interchange terms in our polynomials as shown below, while preserving their root r:

$$\begin{aligned} y_p y_q &\longleftrightarrow N, \\ x_p - 1 &\longleftrightarrow x_q, \\ x_q + 1 &\longleftrightarrow x_p, \\ z_p + 1 &\longleftrightarrow z_q, \\ z_q - 1 &\longleftrightarrow z_p. \end{aligned}$$

With the above replacement rules, we linearize our polynomials as

$$\begin{aligned} f(x_p, x_q, y_p, y_q, z_p, z_q) &:= x_p y_p - x_q, \\ g(x_p, x_q, y_p, y_q, z_p, z_q) &:= y_p z_p - N z_q, \\ h(x_p, x_q, y_p, y_q, z_p, z_q) &:= N x_p z_q - x_q z_p. \end{aligned}$$

By that, all three polynomials have the root r modulo e.

In the following we want to apply the replacement rules to our shift polynomials by using an operator $\mathsf{trans}(\cdot)$ as defined below.

Definition 4. *Let F be a polynomial in the variables $x_p, x_q, y_p, y_q, z_p, z_q$. Then $\mathsf{trans}(F)$ denotes the polynomial, that is obtained by transforming the monomials of F as follows:*

1. *In every monomial replace every $y_p y_q$ by N.*
2. *In every monomial, that has no factor of y_p, replace every x_p by $x_q + 1$ and every z_p by $z_q - 1$.*
3. *In every monomial, that has a factor of y_p, replace every x_q by $x_p - 1$ and every z_q by $z_p + 1$.*

Notice that $\mathsf{trans}(F)$ only has monomials of the form $x_p^a y_p^b z_p^c$ and $x_q^a y_q^b z_q^c$, i.e., variables with subscripts p and q never appear together in one monomial.

As the following lemma shows, polynomials of the form $f^{i_1} y_q^{i_2}$ have a rather nice shape after application of $\mathsf{trans}(\cdot)$.

Lemma 5. *Let $F := f^{i_1} y_q^{i_2}$ with $i_1, i_2 \in \mathbb{N}$ and let $F^* := \mathsf{trans}(F)$. Then the following holds:*

1. *The monomials of F^* are of the form $x_p^a y_p^b$ and $x_q^a y_q^b$.*
2. *The absolute value of the coefficient of $x_p^{i_1} y_p^{i_1 - i_2}$ in F^* is N^{i_2}.*
3. *The absolute value of the coefficient of $x_q^{i_2} y_q^{i_2}$ in F^* is 1.*
4. *If $x_p^a y_p^b$ is a monomial of F^*, then $a \geq b + i_2$.*
5. *If $x_q^a y_q^b$ is a monomial of F^*, then $a \geq b + i_1 - i_2$.*

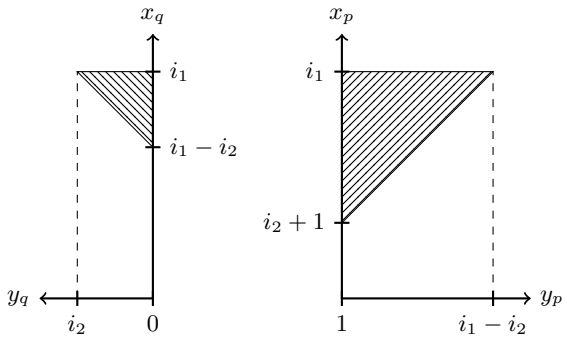

Fig. 6. The Newton polytope of $\mathsf{trans}(f^{i_1} y_q^{i_2})$.

Before we prove Lemma 5, let us give a geometrical interpretation. For $i_1 > i_2 \geq 1$ the Newton polytope of F^* consists of two upper triangles, as shown in Figure 6. Hence, F^* may be written as

$$F^*(x_p, x_q, z_p, z_q) = F_p^*(x_p, y_p) + F_q^*(x_q, y_q),$$

such that the monomials of F_p^* are the elements of the set

$$\left\{ x_p^a y_p^b \mid b > 0, x_p^a y_p^b \text{ is a monomial of } f^{i_1} y_p^{-i_2} \right\},$$

where f has the shape it had before linearization, and similarly the monomials of F_q^* are the elements of the set

$$\left\{ x_q^a y_q^b \mid b \geq 0, x_p^a y_p^b \text{ is a monomial of } f^{i_1} y_p^{i_2 - i_1} \right\}.$$

Proof (Lemma 5). From the equation

$$f^{i_1} = (x_p y_p - x_q)^{i_1} = \sum_{j_1=0}^{i_1} \binom{i_1}{j_1} (x_p y_p)^{i_1 - j_1} (-x_q)^{j_1},$$

we conclude that the monomials of F are of the form $y_p^{i_1-j_1}y_q^{i_2}x_p^{i_1-j_1}x_q^{j_1}$, where $0 \leq j_1 \leq i_1$. By Definition 4, every monomial with $i_1 - j_1 > i_2$ gets transformed via trans as

$$\begin{aligned}
&y_p^{i_1-j_1}y_q^{i_2}x_p^{i_1-j_1}x_q^{j_1}\\
\mapsto &N^{i_2}y_p^{i_1-j_1-i_2}x_p^{i_1-j_1}x_q^{j_1}\\
\mapsto &N^{i_2}y_p^{i_1-j_1-i_2}x_p^{i_1-j_1}(x_p-1)^{j_1}\\
=&N^{i_2}y_p^{i_1-j_1-i_2}x_p^{i_1-j_1}\sum_{j_2=0}^{j_1}\binom{j_1}{j_2}(-1)^{j_2}x_p^{j_1-j_2}\\
=&N^{i_2}y_p^{i_1-j_1-i_2}\sum_{j_2=0}^{j_1}\binom{j_1}{j_2}(-1)^{j_2}x_p^{i_1-j_2}.
\end{aligned}$$

Similarly, every monomial with $i_1 - j_1 \leq i_2$ gets transformed as

$$\begin{aligned}
&y_p^{i_1-j_1}y_q^{i_2}x_p^{i_1-j_1}x_q^{j_1}\\
\mapsto &N^{i_1-j_1}y_q^{i_2-(i_1-j_1)}x_p^{i_1-j_1}x_q^{j_1}\\
\mapsto &N^{i_1-j_1}y_q^{i_2-(i_1-j_1)}(x_q+1)^{i_1-j_1}x_q^{j_1}\\
=&N^{i_1-j_1}y_q^{i_2-(i_1-j_1)}x_q^{j_1}\sum_{j_3=0}^{i_1-j_1}\binom{i_1-j_1}{j_3}x_q^{i_1-j_1-j_3}\\
=&N^{i_1-j_1}y_q^{i_2-(i_1-j_1)}\sum_{j_3=0}^{i_1-j_1}\binom{i_1-j_1}{j_3}x_q^{i_1-j_3}.
\end{aligned}$$

Notice that this already proves the first three statements.

Statements four and five now follow easily. For every monomial $x_p^a y_p^b$ we find values $j_1 = 0, \ldots, i_1$ and $j_2 = 0, \ldots, j_1$, such that $a = i_1 - j_2$ and $b = i_1 - j_1 - i_2$. As this yields the inequality

$$a = i_1 - j_2 \geq i_1 - j_1 = b + i_2,$$

this proves the fourth statement. Similarly, for every every monomial $x_q^a y_q^b$ we find values $j_1 = 0, \ldots, i_1$ and $j_3 = 0, \ldots, i_1 - j_1$, such that $a = i_1 - j_3$ and $b = i_2 - (i_1 - j_1)$. This yields the inequality

$$a = i_1 - j_3 \geq i_1 - (i_1 - j_1) = b + i_1 - i_2$$

and thus concludes the proof of the lemma. □

One can generalize Lemma 5 with a completely analogous proof to the statement of Lemma 6.

Lemma 6. *Let* $F := f^{i_1}g^{i_2}h^{i_3}x_p^{i_4}z_p^{i_5}y_q^{i_6}$ *with* $i_1, \ldots, i_6 \in \mathbb{N}$ *and let* $F^* := \mathsf{trans}(F)$. *Then the following holds:*

1. *The monomials of* F^* *are of the form* $x_p^a y_p^b z_p^c$ *and* $x_q^a y_q^b z_q^c$.
2. *The absolute value of the coefficient of*

$$x_p^{i_1+i_3+i_4} y_p^{i_1+i_2-i_6} z_p^{i_2+i_3+i_5}$$

in F^* *is* $N^{j_1}(N-1)^{i_3}$ *for some* $j_1 \in \mathbb{N}$.
3. *The absolute value of the coefficient of*

$$x_q^{i_1+i_3+i_4} y_q^{i_6} z_q^{i_2+i_3+i_5}$$

in F^* *is* $N^{j_2}(N-1)^{i_3}$ *for some* $j_2 \in \mathbb{N}$.
4. *If* $x_p^a y_p^b z_p^c$ *is a monomial of* F^**, then* $a + c \geq b + i_3 + i_4 + i_5 + i_6$.
5. *If* $x_q^a y_q^b z_q^c$ *is a monomial of* F^**, then* $a + c \geq b + i_1 + i_2 + i_3 - i_6$.

Lemma 6 can be interpreted geometrically analogous to Lemma 5. That is, F^* may be written as

$$F^*(x_p, x_q, y_p, y_q, z_p, z_q) = F_p^*(x_p, y_p, z_p) + F_q^*(x_q, y_q, z_q),$$

such that the monomials of F_p^* are the elements of the set

$$\left\{ x_p^a y_p^b z_p^c \mid b > 0, x_p^a y_p^b z_p^c \text{ is a monomial of } f^{i_1} g^{i_2} h^{i_3} x_p^{i_4} z_p^{i_5} y_p^{-i_6} \right\}, \tag{5}$$

where f, g and h have the shape they had before the linearization, and similarly the monomials of F_q^* are the elements of the set

$$\left\{ x_q^a y_q^b z_q^c \mid b \geq 0, x_p^a y_p^b z_p^c \text{ is a monomial of } f^{i_1} g^{i_2} h^{i_3} (x_p+1)^{i_4} (z_p-1)^{i_5} y_p^{i_6-i_1-i_2} \right\}. \tag{6}$$

Thus, geometrically, the $\mathsf{trans}(\cdot)$ operator creates two copies of the Newton polytope of $f^{i_1} g^{i_2} h^{i_3} x_p^{i_4} z_p^{i_5}$, where one lies in the (x_p, y_p, z_p)-plane and the other one in the (x_q, y_q, z_q)-plane. The larger the exponent of y_q, the larger is the polytope in the (x_q, y_q, z_q)-plane and the smaller is the polytope in the (x_p, y_p, z_p)-plane. In particular, for $i_6 = i_1 + i_2$ the Newton polytope of F^* lies completely in the (x_q, y_q, z_q)-plane, whereas for $i_6 = 0$ it lies completely in the (x_p, y_p, z_p)-plane (except for some monomials $x_q^a y_q^b z_q^c$ with $b = 0$). For $i_6 = (i_1 + i_2)/2$, both components become equally sized. (See also Figure 6.)

Based on this interpretation, we now enhance in the following Proposition 2 our lattice construction from Proposition 1, such that the Newton polytopes of the shift polynomials are equally balanced in both the (x_q, y_q, z_q)-plane and the (x_p, y_p, z_p)-plane.

Proposition 2. *Order the monomials in* $\mathcal{M}$ *according to the* (z_p, x_p, y_p)*-order. Define a lattice basis matrix* $\mathbf{B}$*, in which the* i*-th column corresponds to the monomial*

$$\lambda_{[a,b,c]} := \begin{cases} x_q^a y_q^{b/2} z_q^c, & \text{if } b \text{ is even} \\ x_p^a y_p^{\lceil b/2 \rceil} z_p^c, & \text{if } b \text{ is odd} \end{cases}$$

and the i-th row corresponds to the coefficient vector of

$$p^*_{[a,b,c]} := \mathsf{trans}\left(p_{[a,b,c]} \cdot y_q^{\lfloor b/2 \rfloor}\right)(Xx_p, Xx_q, Yy_p, Yy_q, Xz_p, Xz_q),$$

where $x_p^a y_p^b z_p^c$ is the i-th smallest element in $\mathcal{M}$. Then $\mathbf{B}$ is triangular.

Proof. The proof is similar to that of Proposition 1. We need to show that the i-th polynomial $p^*_{[a,b,c]}$ has with $\lambda_{[a,b,c]}$ exactly one monomial, which is not included in $\mathbf{B}$, before adding $p^*_{[a,b,c]}$ to $\mathbf{B}$. We prove this by induction over i.

Let us first prove the statement for $i = 1$. The smallest element in $\mathcal{M}$ is the monomial $x_p^0 y_p^0 z_p^0 = 1$. Hence, the first column corresponds to $\lambda_{[0,0,0]} = 1$ and the first row corresponds to

$$p^*_{[0,0,0]} = e^{2m} = e^{2m} \cdot \lambda_{[0,0,0]}.$$

As $p^*_{[0,0,0]}$ therefore has with $\lambda_{[0,0,0]}$ exactly one monomial, this proves the statement for $i = 1$.

Now fix an arbitrary $i < |\mathcal{M}|$ and suppose that the statement is true for all $j \leq i$. We show that it then holds for $i+1$. With (5), (6) and Lemma 4 it follows that the $(i+1)$-th polynomial $p^*_{[a,b,c]}$ may be written as

$$p^*_{[a,b,c]}(x_p, x_q, y_p, y_q, z_p, z_q) = p^*_{[a,b,c],p}(x_p, y_p, z_p) + p^*_{[a,b,c],q}(x_q, y_q, z_q),$$

such that:

1. The monomials of $p^*_{[a,b,c],p}$ form a subset of $\mathcal{M}$.
2. The monomials of $p^*_{[a,b,c],q}$ form a subset of $\{x_q^a y_q^b z_q^c \mid x_p^a y_p^b z_p^c \in \mathcal{M}\}$.
3. The leading monomial of $p^*_{[a,b,c],p}$ (according to the (x_p, y_p, z_p)-order) is
$$x_p^a y_p^{b-\lfloor b/2 \rfloor} z_p^c = x_p^a y_p^{\lceil b/2 \rceil} z_p^c.$$
4. The leading monomial of $p^*_{[a,b,c],q}$ (according to a similarly defined (x_q, y_q, z_q)-order) is
$$x_q^a y_q^{b+\lfloor b/2 \rfloor - E_f(a,b,c) - E_g(a,b,c)} z_q^c = x_q^a y_q^{\lfloor b/2 \rfloor} z_q^c.$$

Notice that the equality above follows from the fourth statement in Lemma 3.

Now arguing analogous to the proof of Proposition 1, Proposition 2 easily follows by induction. □

When compared to Proposition 1, the advantage of the lattice construction in Proposition 2 is that we can effectively halve the exponent of Y in the lattice's determinant and by that significantly reduce the determinant's value. One can show (see Remark 3) that the enabling condition (1) now becomes

$$\delta < \frac{5}{56} \approx 0.089. \tag{7}$$

Proposition 2 therefore yields an attack, that already outperforms the Jochemsz-May attack [15].

To further improve the bound on δ to 0.122, Takayasu, Lu and Peng use in [28] basically the lattice construction from Proposition 2, but add extra shifts in the variables y_p and y_q to the lattice, i.e., they include additional polynomials of the form

$$p^*_{[a,b,c,i],q} := \mathsf{trans}\left(p_{[a,b,c]} \cdot y_q^{\lfloor b/2 \rfloor} \cdot y_q^i\right)(Xx_p, Xx_q, Yy_p, Yy_q, Xz_p, Xz_q),$$
$$p^*_{[a,b,c,i],p} := \mathsf{trans}\left(p_{[a,b,c]} \cdot y_q^{\lfloor b/2 \rfloor} \cdot y_p^i\right)(Xx_p, Xx_q, Yy_p, Yy_q, Xz_p, Xz_q).$$

More precisely, whenever adding a polynomial $p^*_{[a,b,c]}$ with $b = a+c$, they include additional rows corresponding to the polynomials

$$\begin{aligned} &p^*_{[a,b,c,1],q}, p^*_{[a,b,c,2],q}, \ldots, p^*_{[a,b,c,\lfloor \tau b \rfloor - \lfloor b/2 \rfloor],q} \\ &p^*_{[a,b,c,1],p}, p^*_{[a,b,c,2],p}, \ldots, p^*_{[a,b,c,\lfloor \tau b \rfloor - \lceil b/2 \rceil],p} \end{aligned} \tag{8}$$

as well as additional columns corresponding to the monomials

$$\begin{aligned} &x_q^a y_q^{\lfloor b/2 \rfloor + 1} z_q^c, x_q^a y_q^{\lfloor b/2 \rfloor + 2} z_q^c, \ldots, x_q^a y_q^{\lfloor \tau b \rfloor} z_q^c, \\ &x_p^a y_p^{\lceil b/2 \rceil + 1} z_p^c, x_p^a y_p^{\lceil b/2 \rceil + 2} z_p^c, \ldots, x_p^a y_p^{\lfloor \tau b \rfloor} z_p^c, \end{aligned} \tag{9}$$

for some parameter $\tau \geq 1/2$, which has to be optimized as a function of δ. Notice that by (4) it follows that none of these monomials are already included in the lattice basis from Proposition 2.

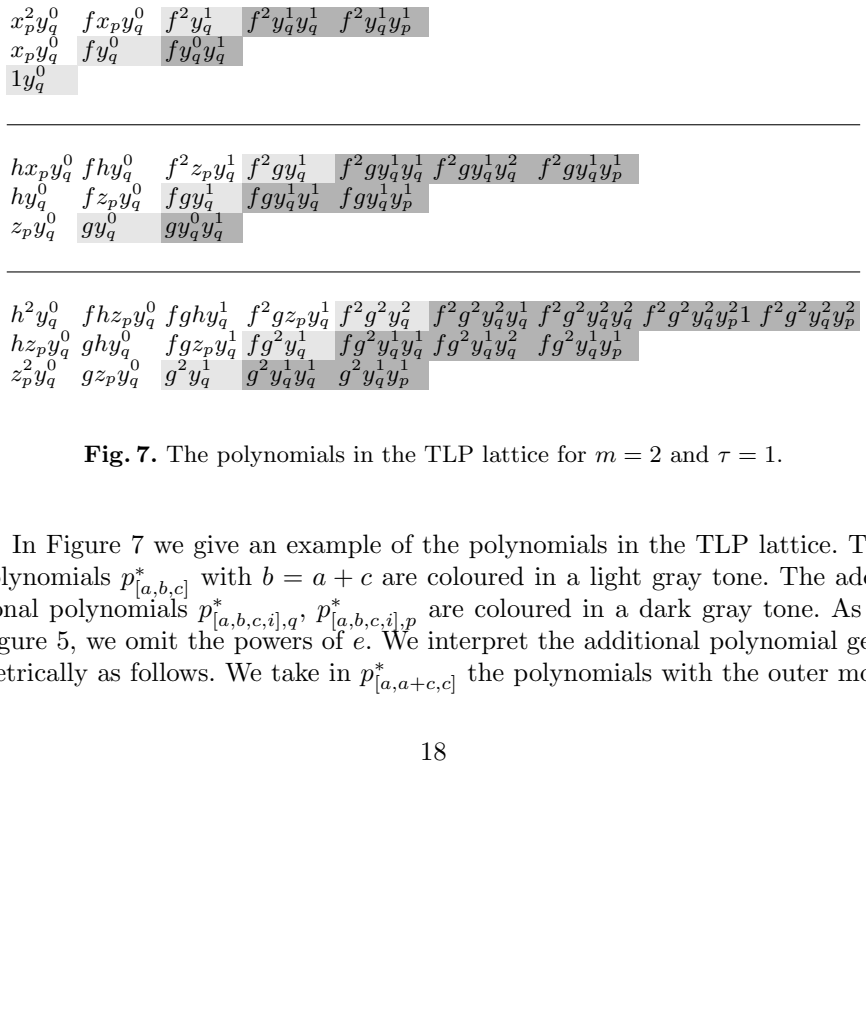

Fig. 7. The polynomials in the TLP lattice for $m = 2$ and $\tau = 1$.

In Figure 7 we give an example of the polynomials in the TLP lattice. The polynomials $p^*_{[a,b,c]}$ with $b = a + c$ are coloured in a light gray tone. The additional polynomials $p^*_{[a,b,c,i],q}$, $p^*_{[a,b,c,i],p}$ are coloured in a dark gray tone. As in Figure 5, we omit the powers of e. We interpret the additional polynomial geometrically as follows. We take in $p^*_{[a,a+c,c]}$ the polynomials with the outer most

Newton polytopes and push these further into the (x_q, y_q, z_q)-plane, respectively the (x_p, y_p, z_p)-plane, by using $p^*_{[a,a+c,c,i],q}$ and $p^*_{[a,a+c,c,i],p}$.

With this interpretation, it is not hard to see that the basis matrix still remains triangular: The polynomial $p^*_{[a,b,c,i],q}$ adds the monomial $x_q^a y_q^{\lfloor b/2 \rfloor + i} z_q^c$ to the lattice basis and $p^*_{[a,b,c,i],p}$ adds $x_p^a y_p^{\lceil b/2 \rceil + i} z_p$. Using this observation, we finally prove the TLP attack.

Theorem 1 (Takayasu, Lu, Peng). *Let $N = pq$ be a sufficiently large RSA modulus, where p and q have the same bit-size. Let $e < \phi(N)$ be a public exponent with $\gcd(e, N-1) = \mathcal{O}(1)$. Suppose the corresponding CRT exponents d_p, d_q are upper bounded by $d_p, d_q \leq N^\delta$, where*

$$\delta < \frac{1}{2} - \frac{1}{\sqrt{7}} \approx 0.122.$$

Given (N, e), we can factor N in polynomial time (under Assumption 1).

Proof. We build a lattice basis matrix $\mathbf{B}$ as in Proposition 2 and add the additional polynomials (8) and monomials (9) as described above. The diagonal elements of $\mathbf{B}$ are products of powers of e, X, Y and (due to statements two and three in Lemma 6) N and $(N-1)$. To reduce the value of the determinant of $\mathbf{B}$, we remove the powers of N and $(N-1)$ as follows. Let

$$\mathbf{B}_{i,i} = e^{E_{1,i}} X^{E_{2,i}} Y^{E_{3,i}} N^{E_{4,i}} (N-1)^{E_{5,i}}$$

denote the i-th diagonal element of $\mathbf{B}$. We replace for every i the value of $\mathbf{B}_{i,i}$ by

$$e^{E_{1,i}} X^{E_{2,i}} Y^{E_{3,i}} \gcd(N-1, e)^{E_{5,i}}$$

and then multiply every other entry in the i-th row of $\mathbf{B}$ by

$$\left(N^{E_{4,i}} \left(\frac{N-1}{\gcd(N-1, e)} \right)^{E_{5,i}} \right)^{-1} \mod e^{2m}.$$

By that, the i-th row still corresponds to a polynomial with the root r modulo e^{2m}.

Notice that we can assume without loss of generality that N is invertible modulo e. If it was not, we could easily obtain a prime factor of N in $\gcd(e, N)$. For $(N-1)$ on the other hand, we of course can not make this assumption and therefore have to use $(N-1)/\gcd(N-1, e)$. Since we have $\gcd(N-1, e) = \mathcal{O}(1)$, we can asymptotically neglect the remaining powers of $\gcd(N-1, e)$ on the diagonal. This allows us to asymptotically calculate the determinant of $\mathbf{B}$ as

$\det \mathbf{B} = e^{s_e} X^{s_x} Y^{s_y}$, where

$$s_e = \sum_{x_p^a y_p^b z_p^c \in \mathcal{M}} E(a,b,c) + \sum_{\substack{x_p^a y_p^b z_p^c \in \mathcal{M}, \\ b=a+c}} 2 \cdot \sum_{i=1}^{\tau b - b/2} E(a,b,c) = \frac{1+5\tau}{3} m^4 + o(m^4),$$

$$s_X = \sum_{x_p^a y_p^b z_p^c \in \mathcal{M}} (a+c) + \sum_{\substack{x_p^a y_p^b z_p^c \in \mathcal{M}, \\ b=a+c}} 2 \cdot \sum_{i=1}^{\tau b - b/2} (a+c) = \frac{7\tau}{3} m^4 + o(m^4),$$

$$s_Y = \sum_{x_p^a y_p^b z_p^c \in \mathcal{M}} \frac{b}{2} + \sum_{\substack{x_p^a y_p^b z_p^c \in \mathcal{M}, \\ b=a+c}} 2 \cdot \sum_{i=1}^{\tau b - b/2} \left(\frac{b}{2} + i \right) = \frac{7\tau^2}{6} m^4 + o(m^4)$$

and

$$E(a,b,c) := 2m - E_f(a,b,c) - E_g(a,b,c) - E_h(a,b,c).$$

Then, calculating the dimension n of the lattice as

$$n = \sum_{x_p^a y_p^b z_p^c \in \mathcal{M}} 1 + \sum_{\substack{x_p^a y_p^b z_p^c \in \mathcal{M}, \\ b=a+c}} 2 \cdot \sum_{i=1}^{\tau b - b/2} 1 = 2\tau m^3 + o(m^3),$$

and plugging in the values $e = N^\alpha$, $X = \Theta(N^{\alpha+\delta-1/2})$ and $Y = \Theta(N^{1/2})$, we find that the enabling condition $\det \mathbf{B} < e^{2mn}$ becomes

$$\alpha \cdot \frac{1+5\tau}{3} m^4 + \left(\alpha + \delta - \frac{1}{2} \right) \cdot \frac{7\tau}{3} m^4 + \frac{1}{2} \cdot \frac{7\tau^2}{6} m^4 < \alpha \cdot 4\tau m^4 + o(m^4). \quad (10)$$

To maximize the bound on δ, we set $\tau := \max\{1 - 2\delta, 1/2\}$, which simplifies the above to

$$\delta < \frac{1}{2} - \sqrt{\frac{\alpha}{7}} + o(1).$$

Notice, the smaller α, the better the bound on δ becomes. Since we have $e < \phi(N)$ and consequently $\alpha < 1$, we can therefore also use the simpler bound

$$\delta < \frac{1}{2} - \frac{1}{\sqrt{7}} + o(1).$$

Consequently, we find for every $\delta < 1/2 - 1/\sqrt{7}$ an m, such that the enabling condition becomes satisfied, which proves the theorem. □

Remark 2. The condition $\gcd(N-1, e) = \mathcal{O}(1)$ does not appear in the original formulation of the theorem in [28]. However, we do not see how to avoid this. If $\gcd(N-1, e)$ becomes large, then we can no longer asymptotically ignore the additional factors on the determinant, and by that obtain a worse bound in the enabling condition. This would imply an inferior bound on δ.

Remark 3. The proof of Theorem 1 can be easily modified to prove the previously mentioned bound (7) of $\delta < 5/56$ for the construction from Proposition 2. If one sets $\tau = 1/2$ in the proof, then no additional polynomials $p^*_{[a,b,c,i],q}$ and $p^*_{[a,b,c,i],p}$ are added to the lattice. Thus, the construction in that case is exactly the same as in Proposition 2. The enabling condition (10) then simplifies to

$$\delta < \frac{3}{8} - \frac{2\alpha}{7} + o(1),$$

which one can further simplify to

$$\delta < \frac{3}{8} - \frac{2}{7} + o(1) = \frac{5}{56} + o(1)$$

by using $\alpha < 1$ as before.

4 Our small CRT-exponent attacks

Our geometrical interpretation of the TLP attack from Section 3 now allows us to easily explain our Partial Key Exposure attack.

As before, let $N = pq$ be an RSA modulus, let $e = N^\alpha$ be a public exponent and let d_p, d_q be the corresponding CRT exponents. We assume that both d_p and d_q are upper bounded by $d_p, d_q \leq N^\beta$ for some $\beta \in \mathbb{R}$. Additionally, we assume that we know integers $\widetilde{d_p}, \widetilde{d_q}, M \approx N^{\beta-\delta}$ (for some $\delta \leq \beta$), such that we can write $d_p = d_p^* M + \widetilde{d_p}$, $d_q = d_q^* M + \widetilde{d_q}$ for some unknown integers $d_p^*, d_q^* \leq N^\delta$. In practice, M might, for instance, be a power of 2 and therefore d_p^*, d_q^* the MSBs of d_p and d_q respectively and $\widetilde{d_p}$, $\widetilde{d_q}$ the LSBs.

In the previous section, we used the equations

$$\begin{aligned} kp - (k-1) &= ed_p, \\ p(\ell - 1) - N\ell &= -ed_q p, \\ k\ell N - (k-1)(l-1) &= e^2 d_p d_q + e(d_p(\ell-1) + d_q(k-1)) \end{aligned}$$

to derive polynomials

$$\begin{aligned} f(x_p, x_q, y_p, y_q, z_p, z_q) &= x_p y_p - x_q, \\ g(x_p, x_q, y_p, y_q, z_p, z_q) &= y_p z_p - N z_q, \\ h(x_p, x_q, y_p, y_q, z_p, z_q) &= N x_p z_q - x_q z_p, \end{aligned}$$

which all have the root $r = (k, k-1, p, q, \ell-1, \ell)$ modulo e. With the additional information given by $\widetilde{d_p}$ and $\widetilde{d_q}$, we can similarly define polynomials

$$\begin{aligned} \widetilde{f}(x_p, x_q, y_p, y_q, z_p, z_q) &:= x_p y_p - x_q - e\widetilde{d_p}, \\ \widetilde{g}(x_p, x_q, y_p, y_q, z_p, z_q) &:= y_p z_p - N z_q + e\widetilde{d_q} y_p, \\ \widetilde{h}(x_p, x_q, y_p, y_q, z_p, z_q) &:= N x_p z_q - x_q z_p - e^2 \widetilde{d_p}\widetilde{d_q} - e\widetilde{d_p} z_p - e\widetilde{d_q} x_q, \end{aligned}$$

which in turn have the root r modulo eM. Notice that for increasing M, the polynomials $\widetilde{f}$, $\widetilde{g}$ and $\widetilde{h}$ are in terms of Coppersmith's method superior to f, g and h, as they have the same small root r modulo a larger modulus. At the same time they are, however, also inferior, since they have more monomials. As we will see below, we therefore obtain our best results, when carefully balancing the use of $\widetilde{f}$, $\widetilde{g}$ and $\widetilde{h}$ with that of f, g and h.

We now use $\widetilde{f}$, $\widetilde{g}$ and $\widetilde{h}$ to build a lattice basis matrix and then apply Coppersmith's method to compute r. We closely follow the construction as described in Proposition 2. However, some modifications are necessary. If we would simply build the lattice exactly as described in Proposition 2, but construct the shift polynomials using $\widetilde{f}$, $\widetilde{g}$ and $\widetilde{h}$ instead of f, g and h, we would not obtain a triangular matrix. For instance, the polynomial $\widetilde{g}$ would add with y_p a new monomial, which does not appear in the lattice from Proposition 2. Overall, we would obtain many additional monomials, as the $\mathsf{trans}(\cdot)$ operator does not work as good with $\widetilde{f}$, $\widetilde{g}$ and $\widetilde{h}$ as it does with f, g and h. Let us illustrate this with an example.

When instantiating the lattice from Proposition 2 with $m = 2$, the shift polynomial $p^*_{[2,2,0]}$ is obtained by multiplying $p_{[2,2,0]} = f^2e^2$ by a factor of $y_q^{\lfloor 2/2 \rfloor} = y_q$ and transforming it using $\mathsf{trans}(\cdot)$ as shown below. (For better readability we omit the factor e^2.)

$$\begin{aligned} f^2 y_q &= (x_p y_p - x_q)^2 y_q \\ &= x_p^2 y_p^2 y_q - 2x_p x_q y_p y_q + x_q^2 y_q \\ &\mapsto N x_p^2 y_p - 2N x_p x_q + x_q^2 y_q \\ &\mapsto N x_p^2 y_p - 2N(x_q + 1)x_q + x_q^2 y_q \\ &= N x_p^2 y_p - 2N x_q^2 - 2N x_q + x_q^2 y_q. \end{aligned}$$

Applying the same transformations to $\widetilde{f}^2$, we obtain

$$\begin{aligned} \widetilde{f}^2 y_q &= (x_p y_p - x_q - e\widetilde{d}_p)^2 y_q \\ &= x_p^2 y_p^2 y_q - 2x_p x_q y_p y_q - 2e\widetilde{d}_p x_p y_p y_q - x_q^2 y_q + 2e\widetilde{d}_p x_q y_q + e^2 \widetilde{d}_p^{\,2} y_q \\ &\mapsto N x_p^2 y_p - 2N x_p x_q - 2Ne\widetilde{d}_p x_p - x_q^2 y_q + 2e\widetilde{d}_p x_q y_q + e^2 \widetilde{d}_p^{\,2} y_q \\ &\mapsto N x_p^2 y_p - 2N(x_q + 1)x_q - 2Ne\widetilde{d}_p(x_q + 1) - x_q^2 y_q + 2e\widetilde{d}_p x_q y_q + e^2 \widetilde{d}_p^{\,2} y_q \\ &= N x_p^2 y_p - 2N x_q^2 - 2N(1 + e\widetilde{d}_p)x_q - 2Ne\widetilde{d}_p - x_q^2 y_q + 2e\widetilde{d}_p x_q y_q + e^2 \widetilde{d}_p^{\,2} y_q. \end{aligned}$$

Comparing the monomials in the variables x_q and y_q of both polynomials in Figure 8, they form a small triangle for the former polynomial, whereas they form a rather large rectangle for the latter.

One can show with a proof analogous to that of Lemma 5 that the shape of the shift polynomials overall becomes more rectangular, when using $\widetilde{f}$, $\widetilde{g}$, $\widetilde{h}$, instead of f, g and h. More precisely, one can show that

$$F^*(x_p, x_q, y_p, y_q, z_p, z_q) := \mathsf{trans}\left(\widetilde{f}^{i_1} \widetilde{g}^{i_2} \widetilde{h}^{i_3} x_p^{i_4} z_p^{i_5} y_q^{i_6}\right)$$

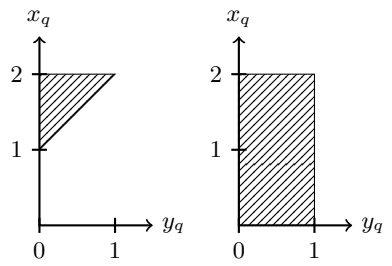

Fig. 8. Parts of the Newton polytopes of $f^2 y_q$ and $\widetilde{f}^2 y_q$ after applying $\mathsf{trans}(\cdot)$.

can be written as

$$F^*(x_p, x_q, y_p, y_q, z_p, z_q) = F_p^*(x_p, y_p, z_p) + F_q^*(x_q, y_q, z_q),$$

such that the monomials of F_p^* form a subset of

$$\left\{ x_p^a y_p^b z_p^b \mid 0 \le a \le i_1 + i_3 + i_4, 0 < b \le i_1 + i_2 - i_6, 0 \le c \le i_2 + i_3 + i_5 \right\}$$

and the monomials of F_q^* form a subset of

$$\left\{ x_q^a y_q^b z_q^b \mid 0 \le a \le i_1 + i_3 + i_4, 0 \le b \le i_6, 0 \le c \le i_2 + i_3 + i_5 \right\}.$$

See Figure 9 for an example.

Additionally, one can show that (as before) the coefficients of

$$x_p^{i_1+i_3+i_4} y_p^{i_1+i_2-i_6} z_p^{i_2+i_3+i_5}$$

and

$$x_q^{i_1+i_3+i_4} y_q^{i_6} z_q^{i_2+i_3+i_5}$$

are non-zero, or more precisely that they are products of powers of N and $(N-1)$. Notice that these monomials correspond to the outer most points in Figure 9, i.e., the points with the largest $\|\cdot\|_1$-norm.

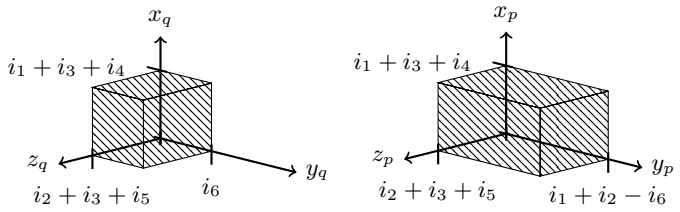

Fig. 9. The effect of $\mathsf{trans}(\cdot)$ on F^*.

As a consequence, we suggest instead of using the set $\mathcal{M}$ for selecting the shift polynomials, to use a different set, which itself has a rectangular shape. For that, we define

$$\widetilde{\mathcal{M}} := \left\{ x_p^a y_p^b z_p^c \mid 0 \le a \le m, 0 \le c \le m, 0 \le b \le 2m \right\}.$$

Notice that the set of tuples (a, b, c) with $x_p^a y_p^b z_p^c \in \widetilde{\mathcal{M}}$ forms a rectangular cuboid of size $m \times 2m \times m$ in $\mathbb{Z}^3$. Also notice that $\mathcal{M} \subseteq \widetilde{\mathcal{M}}$.

We enhance our exponent functions, such that for monomials $x_p^a y_p^b z_p^c \in \widetilde{\mathcal{M}} \setminus \mathcal{M}$ they take the values

$$\begin{aligned} E_f(a,b,c) &:= a, \\ E_g(a,b,c) &:= c, \\ E_h(a,b,c) &:= E_x(a,b,c) := E_z(a,b,c) := 0. \end{aligned}$$

Further, we redefine our shift polynomials as follows:

$$\begin{aligned} \widetilde{p}_{[a,b,c]}(x_p, x_q, y_p, y_q, z_p, z_q) := \widetilde{f}^{E_f(a,b,c)} \cdot \widetilde{g}^{E_g(a,b,c)} \cdot \widetilde{h}^{E_h(a,b,c)} \cdot \\ x_p^{E_x(a,b,c)} \cdot z_p^{E_z(a,b,c)} \cdot \\ (eM)^{2m-(E_f(a,b,c)+E_g(a,b,c)+E_h(a,b,c))}. \end{aligned}$$

Now, to obtain a triangular matrix, our basic idea is to include sufficiently many extra-shifts in y_p and y_q to the lattice, such that for every shift polynomial F^*, every monomial in the cuboids in Figure 9 is included in the basis. We make this strategy more precise in Proposition 3.

Proposition 3. *Order the monomials in $\widetilde{\mathcal{M}}$ according to the (z_p, x_p, y_p)-order. Define a lattice basis matrix $\mathbf{B}$, in which the i-th column corresponds to the monomial*

$$\lambda_{[a,b,c]} := \begin{cases} x_q^a y_q^{b/2} z_q^c, & \text{if } b \text{ is even} \\ x_p^a y_p^{\lceil b/2 \rceil} z_p^c, & \text{if } b \text{ is odd.} \end{cases}$$

where $x_p^a y_p^b z_p^c$ is the i-th smallest element in $\widetilde{\mathcal{M}}$. For $x_p^a y_p^b z_p^c \in \mathcal{M}$, the i-th row of $\mathbf{B}$ corresponds to the coefficient vector of

$$\mathsf{trans}\left(\widetilde{p}_{[a,b,c]} \cdot y_q^{\lfloor b/2 \rfloor} \right) (Xx_p, Xx_q, Yy_p, Yy_q, Xz_p, Xz_q).$$

For $x_p^a y_p^b z_p^c \in \widetilde{\mathcal{M}} \setminus \mathcal{M}$ with even b, the i-th row of $\mathbf{B}$ corresponds to the coefficient vector of

$$\mathsf{trans}\left(\widetilde{p}_{[a,b,c]} \cdot y_q^{\lfloor (a+c)/2 \rfloor} \cdot y_q^{\lceil (b-a-c)/2 \rceil} \right) (Xx_p, Xx_q, Yy_p, Yy_q, Xz_p, Xz_q).$$

For $x_p^a y_p^b z_p^c \in \widetilde{\mathcal{M}} \setminus \mathcal{M}$ with odd b, the i-th row of $\mathbf{B}$ corresponds to the coefficient vector of

$$\mathsf{trans}\left(\widetilde{p}_{[a,b,c]} \cdot y_q^{\lfloor (a+c)/2 \rfloor} \cdot y_p^{\lceil (b-a-c)/2 \rceil} \right) (Xx_p, Xx_q, Yy_p, Yy_q, Xz_p, Xz_q).$$

Then $\mathbf{B}$ is triangular.

As the proof for Proposition 3 is completely analogous to that of Proposition 2, we omit it here.

We are now ready to prove our main theorem.

Theorem 2. *Let $N = pq$ be a sufficiently large RSA modulus, where p and q have the same bit-size. Let $e < \phi(N)$ be a public exponent. Suppose the corresponding CRT exponents d_p, d_q are upper bounded by $d_p, d_q \leq N^\beta$. Write $d_p = d_p^* 2^k + \widetilde{d}_p$, $d_q = d_q^* 2^k + \widetilde{d}_p$, for some $k \in \mathbb{N}$, MSBs $d_p^*, d_q^* \leq N^\delta$ and LSBs $\widetilde{d}_p, \widetilde{d}_q$. If we are given (N, e) and $\widetilde{d}_p, \widetilde{d}_q$, such that*

$$\delta < \frac{1 - 2\beta}{10}$$

and $\gcd(e \cdot 2^k, N - 1) = \mathcal{O}(1)$*, then we can factor N in polynomial time (under Assumption 1).*

Proof. The proof is very similar to that of Theorem 1. We build a lattice basis matrix $\mathbf{B}$ as described in Proposition 3 with $M = 2^k$. As before, we remove the powers of N and $N - 1$ from the diagonal of $\mathbf{B}$ and multiply the other entries in the matix appropriately with the inverses. Notice that as opposed to Theorem 1 here we need the slightly stronger assumption $\gcd(e \cdot 2^k, N - 1) = \mathcal{O}(1)$, as we now have to take inverses modulo eM.

We can asymptotically compute the determinant as $\det \mathbf{B} = (eM)^{s_{eM}} X^{s_X} Y^{s_Y}$, where

$$\begin{aligned}
s_{eM} &= \sum_{x_p^a y_p^b z_p^c \in \widetilde{\mathcal{M}}} (2m - E_f(a,b,c) - E_g(a,b,c) - E_h(a,b,c)) = \frac{7}{3} m^4 + o(m^4), \\
s_X &= \sum_{x_p^a y_p^b z_p^c \in \widetilde{\mathcal{M}}} (a + c) = 2m^4 + o(m^4), \\
s_Y &= \sum_{x_p^a y_p^b z_p^c \in \widetilde{\mathcal{M}}} \frac{b}{2} = m^4 + o(m^4).
\end{aligned}$$

Then, calculating the lattice's dimension as $n = |\widetilde{\mathcal{M}}| = 2m^3$, our enabling condition becomes

$$(\alpha + \beta - \delta) \cdot \frac{7}{3} m^4 + \left(\alpha + \beta - \frac{1}{2}\right) \cdot 2m^4 + \frac{1}{2} \cdot m^4 < (\alpha + \beta - \delta) \cdot 4m^4 + o(m^4).$$

By incorporating $\alpha < 1$ as before, the above simplifies to

$$\delta < \frac{1 - 2\beta}{10} + o(1),$$

which concludes the proof of the theorem. □

In Figure 10 we show for a given value of β, how large of a fraction of key bits is required for the attack to work. That, is on the vertical axis we plot the value $(\beta - \delta)/\beta$.

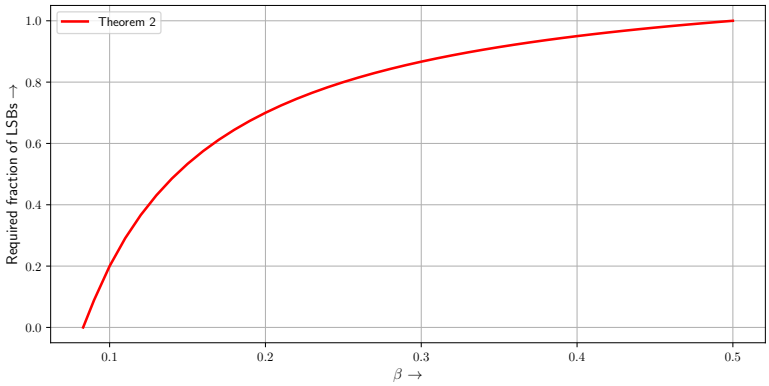

Fig. 10. Required fraction of LSBs to make the attack from Theorem 2 work.

Notice that the graph in Figure 10 has with $(1/2, 1)$ a very natural ending point. The result strongly suggests that for a maximum level of security, full size CRT-exponents must be used – as only then Partial Key Exposure attacks can be prevented. Additionally, it shows that regardless of the key size, we can always factor the modulus, once all key bits are exposed.

Unfortunately, the ending point (1/12,0) on the left side of the graph, however, clearly is non-optimal, as it tells us that for any $\beta > 1/12 \approx 0.083$, at least some key bits have to be exposed to yield the factorization of N. This is contradictory to Theorem 1, by which for any $\beta < 0.122$ no additional key bits are required to factor N.

Intuitively this might be explained with the fact that for $\delta \to \beta$ (i.e., when almost all key bits are unknown) the value eM tends to e. By that, the benefit of using the larger modulus in the lattice construction shrinks more and more as δ grows to β. At a certain point, the inferior shape of the polynomials then outweighs said benefit and therefore gives us an inferior bound. To fill this gap, we propose in the following an alternative lattice construction, inspired by ideas of Aono [1].

4.1 Improved Attack by Linking our First Attack and TLP

The main idea behind the improved construction is to use our lattice from Theorem 2 *together* with the TLP lattice. For that, we define a new set

$$\widetilde{\mathcal{M}}_\sigma := \left\{ x_p^a y_p^b z_p^c \mid 0 \le a \le m, 0 \le c \le m, 0 \le b \le 2\sigma m \right\} \subseteq \widetilde{\mathcal{M}}$$

for some parameter $0 \le \sigma \le 1$, that allows us to interpolate between the TLP lattice and the construction from Proposition 3 and Theorem 2.

When now constructing a lattice exactly as described in Proposition 3, but using the set $\widetilde{\mathcal{M}}_\sigma$ instead of $\widetilde{\mathcal{M}}$, one obtains the same basis matrix $\mathbf{B}_\sigma$, that one

would obtain, when removing all polynomials, which add monomials $x_p^a y_p^b z_p^c$ and $x_q^a y_q^b z_q^c$ with $b > \sigma m$ to the lattice from Proposition 3. Notice that from Figure 9 it follows that the remaining polynomials in $\mathbf{B}_\sigma$ do not have monomials $x_p^a y_p^b z_p^c$ or $x_q^a y_q^b z_q^c$ with $b > \sigma m$. Hence, $\mathbf{B}_\sigma$ is still triangular.

Next, we build another lattice basis matrix $\mathbf{B}_{TLP}$ exactly as described in the TLP attack from Theorem 1, but apply two minor changes:

1. Instead of using the polynomials f, g and h for defining the shift polynomials, we use
$$f^* := Mf, \qquad g^* := Mg, \qquad h^* := Mh.$$
2. We multiply powers of eM to the shift polynomials, instead of powers of e.

Clearly, this does not weaken the TLP attack, as all additional powers of M in the enabling condition cancel out. With these changes, the shift polynomials now have the root r not only modulo e, but also modulo eM. This allows us to combine $\mathbf{B}_\sigma$ and $\mathbf{B}_{TLP}$ as follows.

We remove all polynomials from $\mathbf{B}_{TLP}$ that add monomials $x_p^a y_p^b z_p^c$ and $x_q^a y_q^b z_q^c$ with $b \leq \sigma m$ to the diagonal. After that, we add all polynomials from the matrix $\mathbf{B}_\sigma$ to $\mathbf{B}_{TLP}$. Since in $\mathbf{B}_\sigma$ *all* monomials $x_p^a y_p^b z_p^c$ and $x_q^a y_q^b z_q^c$ with $a \leq m$, $c \leq m$ and $b \leq \sigma m$ appear, it follows that in particular all monomials that we have just removed, reappear in our matrix. Hence, we can rearrange the rows of the newly obtained matrix, such that it is again triangular.

With the above, we thus obtain a triangular lattice basis matrix which nicely incorporates the advantages of the lattice construction from Theorems 1 and 2 at the same time. Similar as in the proofs of both theorems, the enabling condition for the construction becomes

$$e^{s_e} M^{s_M} X^{s_X} Y^{s_Y} < (eM)^{2mn},$$

with analogously defined exponents s_e, s_M, s_X, s_Y and n. (Here we sum over the monomials in $\mathbf{B}_\sigma$ as well as over those in $\mathbf{B}_{TLP}$, except for those that we remove from $\mathbf{B}_{TLP}$.)

For $\sigma \leq \tau$, we have

$$\begin{aligned}
n &= \frac{\sigma^3 + 6\,\tau^3}{3\,\tau^2} m^3 + o(m^3),\\
s_X &= \frac{\sigma^4 + 14\,\tau^4}{6\,\tau^3} m^4 + o(m^4),\\
s_Y &= \frac{3\,\sigma^4\tau + 14\,\tau^5}{12\,\tau^3} m^4 + o(m^4),\\
s_e &= -\frac{\sigma^4 - 4\,\sigma^3\tau - 10\,\tau^4 - 2\,\tau^3}{6\,\tau^3} m^4 + o(m^4),\\
s_M &= -\frac{\sigma^4\tau^2 - 4\,\sigma^3\tau^2 + 6\,\sigma^2\tau^2 - 2\,\sigma^3 + 2\,\sigma\tau^2 - 12\,\tau^3}{3\,\tau^2} m^4 + o(m^4)
\end{aligned}$$

β	0.122	0.123	0.124	0.125	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20	0.21	0.22
$(\beta-\delta)/\beta$	0	0.053	0.084	0.110	0.205	0.332	0.423	0.492	0.549	0.595	0.635	0.0668	0.698	0.723
σ	0	0.328	0.392	0.434	0.548	0.655	0.716	0.757	0.787	0.811	0.830	0.846	0.859	0.869

β	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.30	0.31	0.32	0.33	0.34	0.35	0.36
$(\beta-\delta)/\beta$	0.746	0.767	0.786	0.803	0.819	0.833	0.847	0.859	0.871	0.882	0.892	0.902	0.911	0.919
σ	0.878	0.885	0.891	0.897	0.902	0.907	0.912	0.917	0.922	0.927	0.931	0.935	0.940	0.944

β	0.37	0.38	0.39	0.40	0.41	0.42	0.43	0.44	0.45	0.46	0.47	0.48	0.49	0.50
$(\beta-\delta)/\beta$	0.927	0.934	0.942	0.948	0.955	0.961	0.966	0.972	0.977	0.982	0.987	0.991	0.995	1
σ	0.948	0.952	0.956	0.960	0.964	0.968	0.972	0.976	0.980	0.984	0.988	0.992	0.996	1

Table 1. Values of β, $(\beta-\delta)/\beta$ and σ for our improved lattice construction.

and for $\tau \leq \sigma \leq 2\tau$

$$
\begin{aligned}
n &= -\frac{\sigma^3 - 6\,\sigma^2\tau + 6\,\sigma\tau^2 - 8\,\tau^3}{3\,\tau^2} m^3 + o(m^3),\\
s_X &= -\frac{\sigma^4 - 4\,\sigma^3\tau + 4\,\sigma\tau^3 - 16\,\tau^4}{6\,\tau^3} m^4 + o(m^4),\\
s_Y &= -\frac{3\,\sigma^4\tau - 16\,\sigma^3\tau^2 + 12\,\sigma^2\tau^3 - 16\,\tau^5}{12\,\tau^3} m^4 + o(m^4),\\
s_e &= \frac{\sigma^4 - 8\,\sigma^3\tau + 24\,\sigma^2\tau^2 - 20\,\sigma\tau^3 + 16\,\tau^4 + 2\,\tau^3}{6\,\tau^3} m^4 + o(m^4),\\
s_M &= -\frac{\sigma^4\tau^2 - 4\,\sigma^3\tau^2 + 6\,\sigma^2\tau^2 + 2\,\sigma^3 - 12\,\sigma^2\tau + 14\,\sigma\tau^2 - 16\,\tau^3}{3\,\tau^2} m^4 + o(m^4).
\end{aligned}
$$

Unfortunately, we can not give a closed formula on β and δ as in Theorem 2, because there seems to be no way for analytically maximizing σ. Therefore, we can only present numerical results.

When setting $\tau := \max\{1/2, 1-2\beta\}$ (as in the proof of Theorem 1) and then numerically optimizing σ, we obtain the results shown in Table 1. These results have been used to plot the graph in Figure 2.

Since we reach the lower bound of 0.122, we fully close the gap between Theorems 1 and 2. Notice how the table shows that for $\beta = 0.122$ it is best to use the TLP lattice (i.e., setting $\sigma = 0$) and for $\beta = 0.5$ to use the lattice construction from Proposition 3 (i.e., setting $\sigma = 1$).

5 Experimental Results

The main purpose of our experiments is to verify the validity of Assumption 1.

Although our results theoretically hold in the range $N^{0.122} \leq d_p, d_q \leq N^{0.5}$, we cannot expect to provide experimental data for large d_p, d_q in practice. The reason is that for small exponent CRT-RSA attacks like TLP and our Partial Key Exposure attack the lattice dimension grows as a cubic function in m. Thus, the convergence to the theoretical bounds is quite slow. E.g. for the TLP attack with its theoretical bound $d_p, d_q \leq N^{0.122}$, the original authors provide in [28] practical experiments only up to $N^{0.062}$.

Hence, in order to demonstrate that our attack naturally extends the TLP attack to the Partial Key scenario, we provide some data points with $\beta \geq 0.062$.

We implemented our experiments in SAGE 9.2 using Linux Ubuntu 18.04.4 on a laptop with Intel(R) Core(TM) i7-7920HQ CPU 3.67 GHz. The results are given in Table 2.

Assumption 1 was valid in all experiments. In every run we were able to recover the unknown secrets via Groebner basis computation.

β	Bit-size of N	Bit-size of d_p, d_q	Unknown key-bits	Dimension	LLL Time (sec.)
0.040	1,000	40	2 x 15	53	4
0.040	5,000	200	2 x 80	53	196
0.040	10,000	400	2 x 175	53	1,179
0.065	1,000	65	2 x 20	132	1,242
0.065	5,000	325	2 x 100	132	9,505
0.070	1,000	70	2 x 30	263	51,181
0.100	1,000	100	2 x 30	434	786,423
0.110	1,000	110	2 x 30	434	841,310

Table 2. Experimental results of our Partial Key Exposure attack.

References

1. Aono, Y.: A new lattice construction for partial key exposure attack for RSA. In: International Workshop on Public Key Cryptography. pp. 34–53. Springer (2009)
2. Bernstein, D.J., Chang, Y.A., Cheng, C.M., Chou, L.P., Heninger, N., Lange, T., Van Someren, N.: Factoring RSA keys from certified smart cards: Coppersmith in the wild. In: International Conference on the Theory and Application of Cryptology and Information Security. pp. 341–360. Springer (2013)
3. Bleichenbacher, D., May, A.: New attacks on RSA with small secret CRT-exponents. In: International Workshop on Public Key Cryptography. pp. 1–13. Springer (2006)
4. Blömer, J., May, A.: New partial key exposure attacks on RSA. In: Annual International Cryptology Conference. pp. 27–43. Springer (2003)
5. Boneh, D., Durfee, G.: Cryptanalysis of RSA with private key d less than $N^{0.292}$. IEEE transactions on Information Theory **46**(4), 1339–1349 (2000)
6. Coppersmith, D.: Finding a small root of a bivariate integer equation; factoring with high bits known. In: International Conference on the Theory and Applications of Cryptographic Techniques. pp. 178–189. Springer (1996)
7. Coppersmith, D.: Small solutions to polynomial equations, and low exponent RSA vulnerabilities. Journal of Cryptology **10**(4), 233–260 (1997)
8. Coron, J.S., May, A.: Deterministic polynomial-time equivalence of computing the RSA secret key and factoring. Journal of Cryptology **20**(1), 39–50 (2007)
9. Ernst, M., Jochemsz, E., May, A., De Weger, B.: Partial key exposure attacks on RSA up to full size exponents. In: Annual International Conference on the Theory and Applications of Cryptographic Techniques. pp. 371–386. Springer (2005)
10. Galbraith, S.D., Heneghan, C., McKee, J.F.: Tunable balancing of RSA. In: Australasian Conference on Information Security and Privacy. pp. 280–292. Springer (2005)
11. Heninger, N., Durumeric, Z., Wustrow, E., Halderman, J.A.: Mining your Ps and Qs: Detection of widespread weak keys in network devices. In: 21st USENIX Security Symposium USENIX Security 12. pp. 205–220 (2012)

12. Herrmann, M., May, A.: Attacking power generators using unravelled linearization: When do we output too much? In: International Conference on the Theory and Application of Cryptology and Information Security. pp. 487–504. Springer (2009)
13. Herrmann, M., May, A.: Maximizing small root bounds by linearization and applications to small secret exponent RSA. In: International Workshop on Public Key Cryptography. pp. 53–69. Springer (2010)
14. Howgrave-Graham, N.: Finding small roots of univariate modular equations revisited. In: IMA International Conference on Cryptography and Coding. pp. 131–142. Springer (1997)
15. Jochemsz, E., May, A.: A polynomial time attack on RSA with private CRT-exponents smaller than $N^{0.073}$. In: Annual International Cryptology Conference. pp. 395–411. Springer (2007)
16. Kunihiro, N., Shinohara, N., Izu, T.: A unified framework for small secret exponent attack on RSA. In: International Workshop on Selected Areas in Cryptography. pp. 260–277. Springer (2011)
17. Lenstra, A.K., Lenstra, H.W., Lovász, L.: Factoring polynomials with rational coefficients. Mathematische Annalen **261**(ARTICLE), 515–534 (1982)
18. Lu, Y., Zhang, R., Lin, D.: New partial key exposure attacks on CRT-RSA with large public exponents. In: International Conference on Applied Cryptography and Network Security. pp. 151–162. Springer (2014)
19. Maitra, S., Sarkar, S.: On deterministic polynomial-time equivalence of computing the CRT-RSA secret keys and factoring. Defence Science Journal **62**(2), 122–126 (2012)
20. May, A.: Cryptanalysis of unbalanced RSA with small crt-exponent. In: Yung, M. (ed.) Advances in Cryptology - CRYPTO 2002, 22nd Annual International Cryptology Conference, Santa Barbara, California, USA, August 18-22, 2002, Proceedings. Lecture Notes in Computer Science, vol. 2442, pp. 242–256. Springer (2002)
21. May, A.: New RSA vulnerabilities using lattice reduction methods. Ph.D. thesis, University of Paderborn (2003)
22. May, A.: Computing the RSA secret key is deterministic polynomial time equivalent to factoring. In: Annual International Cryptology Conference. pp. 213–219. Springer (2004)
23. Nemec, M., Sys, M., Svenda, P., Klinec, D., Matyas, V.: The return of Coppersmith's attack: Practical factorization of widely used rsa moduli. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. pp. 1631–1648 (2017)
24. Takayasu, A., Kunihiro, N.: Partial key exposure attacks on RSA: achieving the boneh-durfee bound. In: International Conference on Selected Areas in Cryptography. pp. 345–362. Springer (2014)
25. Takayasu, A., Kunihiro, N.: Partial key exposure attacks on CRT-RSA: better cryptanalysis to full size encryption exponents. In: International Conference on Applied Cryptography and Network Security. pp. 518–537. Springer (2015)
26. Takayasu, A., Kunihiro, N.: How to generalize RSA cryptanalyses. In: Public-Key Cryptography–PKC 2016. pp. 67–97. Springer (2016)
27. Takayasu, A., Lu, Y., Peng, L.: Small CRT-exponent RSA revisited. In: Annual International Conference on the Theory and Applications of Cryptographic Techniques. pp. 130–159. Springer (2017)
28. Takayasu, A., Lu, Y., Peng, L.: Small CRT-exponent RSA revisited. Journal of Cryptology **32**(4), 1337–1382 (2019)
29. Wiener, M.J.: Cryptanalysis of short RSA secret exponents. IEEE Transactions on Information theory **36**(3), 553–558 (1990)