

# Full Cryptanalysis of LPS and Morgenstern Hash Functions

Christophe Petit<sup>1\*</sup>, Kristin Lauter<sup>2</sup> and Jean-Jacques Quisquater<sup>1\*\*</sup>

<sup>1</sup>UCL Crypto Group<sup>\*\*\*</sup>, <sup>2</sup>Microsoft Research.

e-mails:

`christophe.petit@uclouvain.be,klauter@microsoft.com,jjq@uclouvain.be`

**Abstract.** Collisions in the LPS cryptographic hash function of Charles, Goren and Lauter have been found by Zémor and Tillich [15], but it was not clear whether computing preimages was also easy for this hash function. We present a probabilistic polynomial time algorithm solving this problem. Subsequently, we study the Morgenstern hash, an interesting variant of LPS hash, and break this function as well. Our attacks build upon the ideas of Zémor and Tillich but are not straightforward extensions of it. Finally, we discuss fixes for the Morgenstern hash function and other applications of our results.

## 1 Introduction

Hash functions are widely used in cryptographic applications such as commitment schemes, digital signatures schemes, message authentication codes or password encryption. Typically, a hash function is required to be preimage and collision resistant and to have nearly uniform output distribution. Due to the importance of cryptographic hash functions, the SHA family was designed as a NIST standard [2]. However, recently discovered vulnerabilities in SHA-1 [14] prompted NIST to launch a competition for a New Cryptographic Hash Algorithm [1].

The NIST competition is stimulating research on hash functions in the cryptographic community and a lot of new schemes have been recently designed and put forward. Particularly appealing from a theoretical point of view, some of these schemes are *provably secure*, in the sense that their security relates to the hardness of some mathematical problem [8,4,3,12]. A good reduction to a simply formulated mathematical challenge facilitates the evaluation process and increases the confidence once the function has resisted first cryptanalytic attempts. However, it also gives the cryptanalyst a clue to break the scheme, and is especially problematic if the mathematical challenge turns out to be easy.

---

\* Research Fellow of the Belgian Fund for Scientific Research (F.R.S.-FNRS). Part of this work was done while visiting the Security and Cryptography group of the Computer Science Department, UCSD.

\*\* Part of this work was done while visiting MIT (CSAIL-Theory of Computation)

\*\*\* A member of BCRYPT and ECRYPT networks

The LPS hash function proposed by Charles, Goren and Lauter is one of these constructions [3]. It has a particularly elegant design, as the hash computation can be interpreted as a random walk in the optimal expander graphs of Lubotzky, Philips and Sarnak [7]. Finding collisions for this function is finding cycles in the graphs, which also amounts to finding a non-trivial factorization of the identity in terms of some particular elements of a projective group of matrices (a problem we will call the *decomposition problem*). Charles, Goren and Lauter proposed this problem as potentially hard. A major step in the breaking of the LPS hash function has recently been performed by Zémor and Tillich [15] who produced collisions by actually solving this problem.

One of the main contributions of this paper is an efficient algorithm that finds preimages for the LPS hash function. As Zémor and Tillich did, we actually solve the underlying problem which was presumed hard. Both for efficiency considerations and because of these new attacks, it also seemed worth studying the Morgenstern hash function, an interesting variant of the LPS hash relying on different graphs. A second main contribution of this paper is to adapt the Zémor and Tillich attack to Morgenstern hashes, and as an example we give an efficient collision finding algorithm. Combining the ideas of our two algorithms also gives a preimage finding algorithm for Morgenstern hashes that we do not present here due to space limitations.

The paper is organized as follows: in Section 2 we describe the LPS and Morgenstern hash functions; in Section 3 we recall the Zémor and Tillich algorithm; Section 4 presents our preimage algorithm for LPS hashes; in Section 5 we adapt Zémor and Tillich’s algorithm to Morgenstern hashes and in Section 6 we discuss fixes for LPS and Morgenstern hashes, as well as potential applications of our results. In the appendix we give toy examples of our algorithms (1024-bit examples are given in the full version of this paper [10]).

## 2 LPS and Morgenstern hash functions

A *Cayley graph*  $\mathcal{C}_{G,S} = (V, E)$  is a graph constructed from a group  $G$  and a subset  $S$  of  $G$  as follows:  $V$  contains a vertex  $v_g$  associated to each element  $g \in G$ , and  $E$  contains the directed edge  $(v_{g_1}, v_{g_2})$  iff there is some  $s \in S$  such that  $g_2 = g_1s$ . The elements of  $S$  are called the *graph generators*. The graph  $\mathcal{C}_{G,S}$  is  $|S|$ -regular; it is connected iff  $S$  generates  $G$ ; it is undirected iff  $S = S^{-1}$ .

A general construction for a cryptographic hash function from a Cayley graph was introduced by Zémor and Tillich [?,12,13] in the directed case and by Charles, Goren and Lauter [3] in the undirected case. In this paper, we focus on two instances of the undirected graph construction, which we now recall following mainly the description given in [15].

Let  $a := |S| - 1$ . We will define a function  $\pi$  which orders the set of generators (minus one generator to avoid back-tracking). Fix a function  $\pi : \{0, 1 \dots a - 1\} \times S \rightarrow S$  such that for any  $g \in S$  the set  $\pi(\{0, 1 \dots a - 1\} \times \{g\})$  is equal to  $S \setminus \{g^{-1}\}$ . Let  $g_0$  and  $g_{IV}$  be arbitrary fixed elements of  $S$  and  $G$  respectively. The input

message is converted to a base  $a$  number  $x_1 \dots x_k$  and the elements  $g_i = \pi(x_i, g_{i-1})$  are computed recursively. The hashcode of the input message is the product of group elements  $H(x) = g_{1V} g_1 \dots g_k$ .

We will call hash functions constructed following this design strategy *Cayley hashes*. These hash functions have some very interesting properties:

- The girth of the Cayley graph is the length of the smallest cycle, and no two distinct messages of the same length can collide if their length is less than half the girth.
- If the chosen graphs are good expanders (see [6] for precise definitions and applications), the outputs tend to be uniformly distributed, and the convergence to the uniform distribution is fast.
- Differential cryptanalysis (DC), which has been the most successful approach against SHA-1, does not seem to apply to Cayley hashes. Indeed, DC typically activates various portions of the message simultaneously, while in Cayley hashes the bits (or  $k$ -its) are processed one at the time.
- Collision resistance is equivalent to the hardness of a simply-stated representation problem in the corresponding group: namely, this problem is to find a factorization of the identity  $1 = g_1 g_2 \dots g_t$  with  $g_i \in S$  and  $g_i g_{i+1} \neq 1$  for all  $i \in \{1, 2, \dots, t-1\}$ .
- Preimage resistance and second preimage resistance follow from similar problems.

One proposal by Charles, Goren and Lauter [3] is to use the celebrated LPS graphs of Lubotzky, Philips and Sarnak [7] that we now describe. Let  $p$  and  $l$  be primes,  $l$  small and  $p$  large, both  $p$  and  $l$  equal to  $1 \pmod{4}$ , and  $l$  being a quadratic residue modulo  $p$ . To  $l$  and  $p$  is associated an LPS graph  $X_{l,p}$  as follows. Let  $\mathbf{i}$  be an integer such that  $\mathbf{i}^2 \equiv -1 \pmod{p}$ . The vertices of  $X_{l,p}$  are elements in the group  $G = PSL(2, \mathbb{F}_p)$  (*i.e.*  $2 \times 2$  matrices of square determinant, modulo the equivalence relation  $M_1 \sim \lambda M_2, \lambda \in \mathbb{F}_p^*$ ). The set  $S$  is  $S = \{g_j\}_{j=1, \dots, l+1}$ , where

$$g_j = \begin{pmatrix} \alpha_j + \mathbf{i}\beta_j & \gamma_j + \mathbf{i}\delta_j \\ -\gamma_j + \mathbf{i}\delta_j & \alpha_j - \mathbf{i}\beta_j \end{pmatrix}, \quad j = 1, \dots, l+1;$$

and  $(\alpha_j, \beta_j, \gamma_j, \delta_j)$  are all the integer solutions of  $\alpha^2 + \beta^2 + \gamma^2 + \delta^2 = l$ , with  $\alpha > 0$  and  $\beta, \gamma, \delta$  even. The Cayley graph  $X_{l,p} = \mathcal{C}_{G,S}$  is undirected since  $S$  is stable under inversion.

The choice of LPS graphs was very appealing : they are Ramanujan (*i.e.*, they have optimal expansion properties asymptotically [7,6]); they have no short cycles, and computing the resulting hash functions turned out to be quite efficient compared to other provable hashes [?]. Unfortunately, it turns out that LPS hash function is neither collision nor preimage resistant (see Sections 3 and 4 below).

For efficiency reasons, we recently considered the use of Morgenstern graphs to replace LPS graphs in Charles-Goren-Lauter's construction [?]. Morgenstern's Ramanujan graphs [9] generalize LPS graphs from an odd prime  $p \equiv 1 \pmod{4}$  to

any power of any prime  $q$ . More specifically, we suggested the use of Morgenstern graphs with  $q = 2^k$ , that we now describe.

Let  $q$  be a power of 2 and  $f(x) = x^2 + x + \epsilon$  irreducible in  $\mathbb{F}_q[x]$ . Let  $p(x) \in \mathbb{F}_q[x]$  be irreducible of even degree  $n = 2d$  and let  $\mathbb{F}_{q^n}$  be represented by  $\mathbb{F}_q[x]/(p(x))$ . The vertices of the Morgenstern graph  $I_q$  are elements of  $G = PSL_2(\mathbb{F}_{q^n})$  (i.e.  $2 \times 2$  matrices modulo the equivalence relation  $M_1 \sim \lambda M_2, \lambda \in \mathbb{F}_{q^n}^*$ ). Let  $\mathbf{i} \in \mathbb{F}_{q^n}$  be a root of  $f(x)$ . The set  $S$  is taken to be  $S = \{g_j\}_{j=1, \dots, q+1}$ , where

$$g_j = \begin{pmatrix} 1 & \gamma_j + \delta_j \mathbf{i} \\ (\gamma_j + \delta_j \mathbf{i} + \delta_j)x & 1 \end{pmatrix}, \quad j = 1, \dots, q+1;$$

where  $\gamma_j, \delta_j \in \mathbb{F}_q$  are all the  $q+1$  solutions in  $\mathbb{F}_q$  for  $\gamma_j^2 + \gamma_j \delta_j + \delta_j^2 \epsilon = 1$ . The Cayley graphs  $I_q = \mathcal{C}_{G,S}$  are also undirected as each  $g_j$  has order 2.

An interesting property of Morgenstern hashes compared to LPS hashes is that arithmetic is done in fields that are extensions of  $\mathbb{F}_2$  rather than in finite prime fields, potentially leading to faster hashes for some architectures. The total break of LPS hashes leads to the question of whether similar attacks can be found for Morgenstern hashes. This is indeed the case, and as an example we give a collision-finding attack for  $q = 2$  in Section 5.

### 3 Zémor and Tillich algorithm

As our new attacks will build upon it, we now briefly recall Zémor and Tillich's algorithm that finds collisions for LPS hashes [15]. The algorithm lifts the graph generators and the representation problem from  $PSL(2, \mathbb{F}_p)$  to an appropriate subset  $\Omega$  of  $SL(2, \mathbb{Z}[i])$  (in this section and the next one,  $i$  is the complex imaginary number satisfying  $i^2 + 1 = 0$  while  $\mathbf{i}$  is a solution to  $\mathbf{i}^2 + 1 \equiv 0 \pmod{p}$ ). The relevant set is

$$\Omega = \left\{ \begin{pmatrix} a + bi & c + di \\ -c + di & a - bi \end{pmatrix} \mid (a, b, c, d) \in E_e \text{ for some integer } e > 0 \right\}$$

where  $E_e$  is the set of 4-tuples  $(a, b, c, d) \in \mathbb{Z}^4$  such that

$$\begin{cases} a^2 + b^2 + c^2 + d^2 = l^e \\ a > 0, a \equiv 1 \pmod{2} \\ b \equiv c \equiv d \equiv 0 \pmod{2}. \end{cases}$$

We will call the first of these equations describing  $E_e$  the *norm equation*, as the left-hand side of this equation is the norm of the quaternion corresponding to the quadruplet  $(a, b, c, d)$  (see [7]). The set  $\Omega$  has two important properties: first, any element of  $\Omega$  admits a unique factorization in terms of the lifts of the graph generators, and second, there exists a multiplicative homomorphism from  $\Omega$  to  $PSL(2, \mathbb{F}_p)$  that allows translation of this factorization back to  $PSL(2, \mathbb{F}_p)$ .

In their exposition, Zémor and Tillich decompose their attack into three steps. The first step (lifting the decomposition problem to  $SL(2, \mathbb{Z}[i])$ ) amounts to finding integers  $a, b, c, d$  and  $\lambda$  satisfying the following conditions:

$$\begin{cases} (a, b, c, d) \in E_e \\ (a, b, c, d) \text{ not divisible by } 1 \\ (a, b, c, d) \equiv \lambda(1, 0, 0, 0) \pmod{p}. \end{cases}$$

Putting every congruence condition into the norm equation leads to a diophantine equation that was solved by Zémor and Tillich in their paper. The second step of the attack is to factorize the lifted element  $I'$  of  $\Omega$  into products of lifted generators  $g'_j$ ,  $j = 1..l+1$ . We know this factorization is unique and has size  $e$ , so let us write it  $I' = g'_{j_1} g'_{j_2} \dots g'_{j_e}$ . Multiplying on the right by a lifted generator  $g'$  gives a matrix that is divisible by  $l$  if and only if  $g' = (g'_{j_e})^{-1}$ , so by trying each of the graph generators we get the last factor, and we then proceed recursively. The final step is to transpose the factorization of  $I'$  in  $\Omega$  into a factorization of the identity in  $PSL(2, \mathbb{F}_p)$ , but using the homomorphism from  $\Omega$  to  $PSL(2, \mathbb{F}_p)$ , this last step is trivial. For details on the attack we refer to [15].

#### 4 Finding preimages for LPS hashes

Suppose we are given a matrix  $M = \begin{pmatrix} M_1 & M_2 \\ M_3 & M_4 \end{pmatrix} \in PSL(2, \mathbb{F}_p)$  which has square determinant, and we are asked to find a preimage, that is a factorization of it with the graph generators. By solving two linear equations in  $\mathbb{F}_p$  we can write it in the form

$$M = \begin{pmatrix} A + B\mathbf{i} & C + D\mathbf{i} \\ -C + D\mathbf{i} & A - B\mathbf{i} \end{pmatrix}.$$

Our algorithm follows along the lines of Zémor and Tillich's. We first lift the problem from  $PSL(2, \mathbb{F}_p)$  to the set  $\Omega$  defined above, then factorize in  $\Omega$  and finally come back to  $PSL(2, \mathbb{F}_p)$ . The only difference will be in the first step. Lifting the representation problem now amounts to finding integers  $a, b, c, d$  and  $\lambda$  satisfying the following conditions:

$$\begin{cases} (a, b, c, d) \in E_e \\ (a, b, c, d) \text{ not divisible by } 1 \\ (a, b, c, d) \equiv \lambda(A, B, C, D) \pmod{p}. \end{cases}$$

We write  $a = A\lambda + wp$ ,  $b = B\lambda + xp$ ,  $c = C\lambda + yp$  and  $d = D\lambda + zp$  with  $w, x, y, z \in \mathbb{Z}$ . For convenience we choose  $e$  even, that is  $e = 2k$  for  $k$  an integer. The norm equation becomes

$$(A\lambda + wp)^2 + (B\lambda + xp)^2 + (C\lambda + yp)^2 + (D\lambda + zp)^2 = l^{2k}. \quad (1)$$

In the case  $B = C = D = 0$  the norm equation is  $(A\lambda + wp)^2 + (xp)^2 + (yp)^2 + (zp)^2 = l^{2k}$  and was solved by Zémor and Tillich as follows: choose

$$A\lambda + wp = l^k + mp^2$$

for small  $m$  and appropriate  $k$ , hence the equation is already satisfied modulo  $p^2$ . Simplifying by  $p^2$  we get a quadratic diophantine equation of type  $x^2 + y^2 + z^2 = m(l^k - mp^2)$  which Zémor and Tillich show has a solution either for  $m = 1$  or for  $m = 2$ . In Equation 1, when  $B, C, D$  are non-zero we cannot divide by  $p^2$  because of the term  $2p(wA + xB + yC + zD)\lambda$ . Since we do not, the coefficients of degree-2 terms are huge (at least  $p$ ), and the equation is at first sight very hard to solve.

We overcome this difficulty with a new idea. In the remainder of this section, we will solve the preimage problem for diagonal matrices with  $A$  and/or  $B$  non-zero, and then we will write any matrix as a product of four diagonal matrices and up to four graph generators. Altogether this leads to an efficient probabilistic algorithm that finds preimages of the LPS hash function.

*Preimages for diagonal matrices* Now we show how to find a factorization of a matrix

$$M = \begin{pmatrix} A + Bi & \\ & A - Bi \end{pmatrix}$$

such that  $A^2 + B^2$  is a square modulo  $p$ . Write  $y = 2y'$  and  $z = 2z'$  where  $y', z'$  are integers. We need to find integer solutions to

$$\begin{cases} (A\lambda + wp)^2 + (B\lambda + xp)^2 + 4p^2(y'^2 + z'^2) = l^{2k} \\ A\lambda + wp \equiv 1 \pmod{2} \\ B\lambda + xp \equiv 0 \pmod{2} \end{cases}$$

Fix  $k = \lceil \log_l(8p^2) \rceil$ . As  $A^2 + B^2$  is a square, there are exactly two values for  $\lambda$  in  $\{0, 1, \dots, p-1\}$  satisfying the norm equation modulo  $p$ :

$$(A^2 + B^2)\lambda^2 = l^{2k} \pmod{p}.$$

Choose either of them, and let  $m := (l^{2k} - (A^2 + B^2)\lambda^2)/p$ . Our strategy will be to pick random solutions to

$$\begin{cases} l^{2k} - (A\lambda + wp)^2 - (B\lambda + xp)^2 \equiv 0 \pmod{p^2} \\ A\lambda + wp \equiv 1 \pmod{2} \\ B\lambda + xp \equiv 0 \pmod{2} \end{cases}$$

until the equation

$$4(y'^2 + z'^2) = n$$

has solutions, where

$$n := (l^{2k} - (A\lambda + wp)^2 - (B\lambda + xp)^2) / p^2.$$

A random solution to the congruence system is computed as follows: until you get  $x$  with the correct parity, pick a random  $w \in \{0, 1, \dots, p-1\}$  with the right parity and compute  $x = \frac{m}{2\lambda B} - \frac{A}{B}w \pmod{p}$ . By the way  $k$ ,  $x$  and  $w$  are chosen we are guaranteed that  $n > 0$  so the equation  $4(y'^2 + z'^2) = n$  has solution if and

only if 4 divides  $n$  and all prime factors of  $n$  congruent to 3 modulo 4 appear an even number of times in the factorization of  $n$ . To avoid the factorization of  $n$  in the algorithm, we will actually strengthen this condition to  $n$  being equal to 4 times a prime congruent to 1 modulo 4. When it has solutions, the equation  $4(y'^2 + z'^2) = n$  is easily solved with the Euclidean algorithm, as recalled in [15]. After lifting the problem to  $SL(2, \mathbb{Z}[i])$  the second and third steps of the algorithm are the same as in Zémor-Tillich algorithm. So we are done with the factorization of diagonal matrices.

*Reduction to the diagonal case* Now we show how to decompose any matrix  $M \in PSL(2, \mathbb{F}_p)$  into a product of diagonal matrices and graph generators. We may additionally assume that all the entries of  $M$  are nonzero: if they are not, just multiply  $M$  by  $gg^{-1}$  for some adequate  $g$  in  $S$ , and consider the factorization of  $g^{-1}M$ . We will show how to find  $(\lambda, \alpha, \omega, \beta_1, \beta_2)$  with the last four being squares, such that

$$\begin{pmatrix} M_1 & M_2 \\ M_3 & M_4 \end{pmatrix} = \lambda \begin{pmatrix} 1 & 0 \\ 0 & \alpha \end{pmatrix} \begin{pmatrix} f_1 & f_2 \\ f_3 & f_4 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \omega \end{pmatrix} = \lambda \begin{pmatrix} f_1 & \omega f_2 \\ \alpha f_3 & \alpha \omega f_4 \end{pmatrix} \quad (2)$$

and

$$\begin{aligned} \begin{pmatrix} f_1 & f_2 \\ f_3 & f_4 \end{pmatrix} &= \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \beta_1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \beta_2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 - 4\beta_1 - 4\beta_2 - 4\beta_1\beta_2 & 2 - 8\beta_1 + 2\beta_2 + 2\beta_1\beta_2 \\ -2 - 2\beta_1 + 8\beta_2 - 2\beta_1\beta_2 & -4 - 4\beta_1 - 4\beta_2 + \beta_1\beta_2 \end{pmatrix}. \end{aligned}$$

**Lemma 1.** *Matrix equation (2) is equivalent to the following system:*

$$\begin{cases} M_2 M_3 f_1 f_4 - M_1 M_4 f_2 f_3 = 0 \\ \alpha M_1 f_3 - M_3 f_1 = 0 \\ \omega M_3 f_4 - M_4 f_3 = 0 \\ \lambda f_1 - M_1 = 0 \end{cases} \quad (3)$$

Proof : ( $\Rightarrow$ ) Fourth equation is entry (1,1) of the matrix equation. Third equation is entry (2,1) times  $M_1$  minus entry (1,1) times  $M_3$ . Second equation is entry (1,2) times  $M_1$  minus entry (1,1) times  $M_2$ . First equation is entry (1,1) times entry (2,2) times  $M_2 M_3$  minus entry (1,2) times entry (2,1) times  $M_1 M_4$ .

( $\Leftarrow$ ) Last equation is  $M_1 = \lambda f_1$  that is entry (1,1). We have  $M_2 = \frac{M_1 M_4 f_2 f_3}{M_3 f_1 f_4}$  by first equation so  $M_2 = f_2 \frac{M_4 f_3}{M_3 f_4} \frac{M_1}{f_1} = f_2 \omega \lambda$  by third and fourth equation, that is entry (1,2). We have  $M_3 = \frac{\alpha M_1 f_3}{f_1} = \alpha \lambda f_3$  by second then fourth equation, that is entry (2,1). We have  $M_4 = \omega M_3 \frac{f_4}{f_3}$  by third equation, so using the already proved entry (2,1) we have  $M_4 = \omega \alpha \lambda f_3 \frac{f_4}{f_3} = \omega f_4 \alpha \lambda$  that is entry (2,2).  $\square$

In the system of equations (3), the first equation only involves  $\beta_1$  and  $\beta_2$  while the other equations are linear once  $\beta_1$  and  $\beta_2$  are fixed. So we can concentrate

on solving the first equation, which is quadratic in both  $\beta_1$  and  $\beta_2$ :

$$\begin{aligned} M_2M_3f_1f_4 - M_1M_4f_2f_3 &= 4(M_2M_3 - M_1M_4)(-\beta_1^2 + 3\beta_1 + 4)\beta_2^2 \\ &\quad + (M_2M_3(12\beta_1^2 + 49\beta_1 + 12) + M_1M_4(-12\beta_1^2 + 76\beta_1 - 12))\beta_2 \\ &\quad + 4(M_2M_3 - M_1M_4)(4\beta_1^2 + 3\beta_1 - 1). \end{aligned}$$

Our algorithm then proceeds as follows:

1. Pick a random  $\beta_1$  which is a square.
2. Compute the discriminant of the quadratic equation in  $\beta_2, \beta_1$ . If it is not a square, go back to 1.
3. Solve the quadratic equation. If none of the roots is a square, go back to 1. Else, assign a quadratic root to  $\beta_2$ .
4. Compute  $f_1, f_2, f_3, f_4$ .
5. Solve  $\alpha M_1f_3 - M_3f_1 = 0$  to get  $\alpha$ . If  $\alpha$  is not a square, go back to 1.
6. Solve  $\omega M_3f_4 - M_4f_3 = 0$  to get  $\omega$ . If  $\omega$  is not a square, go back to 1.

This concludes the exposition of our algorithm.

*Runtime analysis* First consider the algorithm for diagonal matrices. Assuming  $n$  behaves “as a random number” then according to the prime number theorem we will need  $\mathcal{O}(\log n) = \mathcal{O}(\log p)$  trials before getting one  $n$  of the correct form. For each trial, the most expensive computation is a primality test, which can be done in polynomial time (in our implementation, we actually use the probabilistic function `mpz_probab_prime_p` of GNU MP). So the algorithm for diagonal matrices is probabilistic polynomial time. In the reduction algorithm, the probability for a random number to be a square modulo  $p$  being one half, we estimate that a solution  $(\lambda, \alpha, \omega, \beta_1, \beta_2)$  with the last four being squares can be found in about  $2^4$  trials. Consequently, the whole algorithm is probabilistic polynomial time. Our implementation using GNU MP finds preimages in less than 2 minutes for 1024-bit parameters on an Intel Pentium M processor 1.73GHz.

## 5 Collisions for the Morgenstern hash function

Now we show how to adapt Zémor and Tillich’s algorithm for finding collisions in Morgenstern hashes when  $q = 2$ . Our algorithm lifts the representation problem from  $SL(2, \mathbb{F}_{2^n})$  to a subset  $\Omega$  of  $SL(2, \mathbb{A})$  where  $\mathbb{A} = \mathbb{F}_2[x, y]/(y^2 + y + 1)$  (in this section,  $i$  will denote a root of  $i^2 + i + 1 = 0$  in  $\mathbb{A}$  while  $\mathbf{i}$  is a root of the same equation in  $\mathbb{F}_{2^n}$ ). The relevant set is

$$\Omega = \left\{ \begin{pmatrix} a + bi & c + di \\ x(c + di + d) & a + bi + b \end{pmatrix} \mid (a, b, c, d) \in E_e \text{ for some integer } e > 0 \right\}$$

where  $E_e$  is the set of 4-tuples  $(a, b, c, d) \in \mathbb{F}_2[x]$  such that

$$\begin{cases} (a^2 + b^2 + ab) + (c^2 + d^2 + cd)x = (1 + x)^e \\ a \equiv 1 \pmod{x} \\ b \equiv 0 \pmod{x}. \end{cases}$$



Again call the first of these equations the *norm equation*. We point out that in this section, small letters  $a, b, c, d, i, p$  are polynomials in  $x$  over  $\mathbb{F}_2$ , while capitalized letters will be used for elements of the field  $\mathbb{F}_{2^n}$ . By [9], corollary 5.4 and 5.7, if we restrict  $E_e$  to tuples  $(a, b, c, d)$  not divisible by  $(1+x)$ , the elements of  $\Omega$  have a unique factorization in terms of the lifts of the graph generators:

$$g'_0 = \begin{pmatrix} 1 & 1+i \\ ix & 1 \end{pmatrix}, \quad g'_1 = \begin{pmatrix} 1 & 1 \\ x & 1 \end{pmatrix}, \quad g'_2 = \begin{pmatrix} 1 & i \\ (1+i)x & 1 \end{pmatrix}.$$

Moreover, the “reduction modulo  $p$ ”  $(a, b, c, d) \rightarrow (A, B, C, D) = (a, b, c, d) \bmod p$  gives a homomorphism from  $\Omega$  to  $SL(2, \mathbb{F}_{2^n})$ :

$$\begin{pmatrix} a+bi & c+di \\ x(c+di+d) & a+bi+b \end{pmatrix} \rightarrow \begin{pmatrix} A+B\mathbf{i} & C+D\mathbf{i} \\ x(C+D\mathbf{i}+D) & A+B\mathbf{i}+B \end{pmatrix}.$$

From this it is now clear how the second and third steps of Zémor and Tillich algorithm will work for Morgenstern hashes, so we now give details for first step. This amounts to lifting the representation problem, that is finding  $a, b, c, d, \lambda \in \mathbb{F}_{2^n}$  satisfying the following conditions:

$$\begin{cases} (a, b, c, d) \in E_e \\ (a, b, c, d) \text{ not divisible by } x+1 \\ (a, b, c, d) \equiv \lambda(1, 0, 0, 0) \bmod p. \end{cases}$$

Write  $b = xpb'$ ,  $c = pc'$ ,  $d = pd'$  for  $b', c', d' \in \mathbb{F}_2[x]$  and arbitrarily choose  $e = 2k$  and  $a = (1+x)^k + xpm$ , with  $k \in \mathbb{Z}$  and  $m \in \mathbb{F}_2[x]$  still to be determined. Note that such an  $a$  satisfies  $a \equiv 1 \bmod x$ . The norm equation becomes

$$x^2p^2m^2 + x^2p^2b'^2 + xpb'((1+x)^k + xpm) + xp^2(c'^2 + d'^2 + c'd') = 0.$$

Simplifying by  $xp$  we get

$$xpm^2 + xpb'^2 + b'((1+x)^k + xpm) + p(c'^2 + d'^2 + c'd') = 0.$$

Reducing this equation modulo  $p$  we get  $b'((1+x)^k + xpm) \equiv 0$  which implies  $b' = pb''$  for some  $b'' \in \mathbb{F}_p$ . The norm equation becomes

$$xpm^2 + xp^3b''^2 + pb''((1+x)^k + xpm) + p(c'^2 + d'^2 + c'd') = 0.$$

Simplifying again by  $p$  we get

$$c'^2 + d'^2 + c'd' = n(b'', m, k) := xm^2 + xp^2b''^2 + b''(1+x)^k + b''xpm.$$

Our approach for step 1 will be to generate random  $m$  and  $b''$  (with  $x+1 \nmid b''$ ) until the equation  $c'^2 + d'^2 + c'd' = n(b'', m, k)$  has solutions, then to solve this equation for  $c', d'$ . As will be clear later, the equation has a solution if and only if all the irreducible factors of  $n$  are of even degree. So in particular

- We will choose  $b'' = b^{(3)}x + 1$  for some  $b^{(3)} \in \mathbb{F}_2[x]$  to avoid an  $x$  factor.

- As the term  $xp^2b''^2$  is of odd degree, we will make another term of higher even degree, with the following strategy:
  - Choose  $b''$  and  $m$  randomly of degree equal to or less than  $R$ .
  - Choose  $k = 2 \deg(p) + \deg(b'') + 2 + (\deg(b'') + \epsilon)$  where  $\epsilon = 0$  if  $\deg(b'')$  is even and  $\epsilon = 1$  if  $\deg(b'')$  is odd.

If  $R$  is large enough we get an  $n$  with the desired property after sufficiently many random trials on  $b''$  and  $m$ . In our implementation, we chose  $R = 10$  which is more than enough for 1024-bit parameters. It remains to show how to solve the equation  $c'^2 + d'^2 + c'd' = n$  and to explain the condition on the degrees of irreducible factors of  $n$ . We begin with the solution of the equation.

*Solving  $c^2 + d^2 + cd = n$ .* It is enough to have an algorithm solving it when  $n$  is irreducible. Indeed, if  $c_1^2 + d_1^2 + c_1d_1 = n_1$  and  $c_2^2 + d_2^2 + c_2d_2 = n_2$  then  $(c_3, d_3) = (c_1c_2 + d_1d_2, c_1d_2 + c_2d_1 + d_1d_2)$  satisfies  $c_3^2 + d_3^2 + c_3d_3 = n_1n_2$ . So suppose  $n$  is irreducible of even degree.

We describe a continued fraction algorithm for polynomials over  $\mathbb{F}_2$  and then we use it to solve the equation. For a fraction  $\xi = \frac{P}{Q}$  where  $P$  and  $Q$  are polynomials, let  $P = a_0Q + r_0$  where  $\deg r_0 < \deg Q$ . Let  $Q = a_1r_0 + r_1$  with  $\deg r_1 < \deg r_0$ , then recursively for  $i = 2, \dots$ , define  $r_{i-2} = a_i r_{i-1} + r_i$  with  $\deg r_i < \deg r_{i-1}$ . (This is the Euclidean algorithm applied to the ring  $\mathbb{F}_2[x]$ .) Define  $p_0 = a_0$ ,  $q_0 = 1$ ,  $p_1 = a_0a_1 + 1$ ,  $q_1 = a_1$ , and then recursively  $p_i = a_i p_{i-1} + p_{i-2}$  and  $q_i = a_i q_{i-1} + q_{i-2}$ . (The fraction  $p_i/q_i$  is the  $i^{\text{th}}$  truncated continued fraction of  $P/Q$ .) We see recursively that  $q_i p_{i-1} + q_{i-1} p_i = 1$ , so  $\frac{p_i}{q_i} + \frac{p_{i-1}}{q_{i-1}} = \frac{1}{q_{i-1}q_i}$  and

$$\frac{P}{Q} = a_0 + \sum_{i=0}^n \frac{1}{q_{i+1}q_i}$$

where  $n$  is the first  $i$  such that  $p_i/q_i = P/Q$ . Define a “norm”  $v$  on quotients of polynomials as follows:  $v\left(\frac{a}{b}\right) = \deg a - \deg b$  if  $a, b \neq 0$ ,  $v\left(\frac{a}{b}\right) = 0$  if  $b = 0$ , and  $v\left(\frac{a}{b}\right) = -\infty$  if  $a = 0$ . Note that  $v(q_{i+1}) \geq v(q_i)$ ,  $v(p_{i+1}) \geq v(p_i)$ , and that

$$v\left(\frac{P}{Q} + a_0 + \sum_{i=0}^{n'-1} \frac{1}{q_{i+1}q_i}\right) = v\left(\sum_{i=n'}^n \frac{1}{q_{i+1}q_i}\right) \leq -v(q_{n'+1}) - v(q_{n'})$$

As  $n$  has even degree, we can compute  $\alpha \in \mathbb{F}_2[x]$  such that  $\alpha^2 + \alpha + 1 \equiv 0 \pmod{n}$  (see next paragraph). We apply a continued fraction expansion to  $\xi = \frac{\alpha}{n}$  and let  $p_i/q_i$  be the successive approximations. Let  $j$  be such that

$$v(q_j) \leq \frac{v(n)}{2} \leq v(q_{j+1}).$$

We have

$$q_j^2 + (q_j\alpha + p_jn)^2 + q_j(q_j\alpha + p_jn) \equiv q_j^2 + q_j^2\alpha^2 + q_j^2\alpha \equiv 0 \pmod{n}.$$

On the other hand, as

$$\deg(q_j\alpha + p_jn) = v(n) + v(q_j) + v\left(\xi + \frac{p_j}{q_j}\right) \leq v(n) + v(q_j) - v(q_j) - v(q_{j+1}) \leq v(n)/2 = \deg(n)/2$$

we have

$$v(q_j^2 + (q_j\alpha + p_jn)^2 + q_i(q_j\alpha + p_jn)) \leq 2 \max(\deg(q_j), \deg(q_j\alpha + p_jn)) \leq \deg n.$$

Consequently,

$$q_j^2 + (q_j\alpha + p_jn)^2 + q_j(q_j\alpha + p_jn) = n$$

and  $(c, d) = (q_j, q_j\alpha + p_jn)$  is a solution to  $c^2 + d^2 + cd = n$ .

*Solutions to  $\alpha^2 + \alpha + 1 \equiv 0 \pmod n$ .* As the map  $x \rightarrow x^2 + x$  is linear in  $\mathbb{F}_2$ , solutions to this equation, if there are any, are found easily by writing down then solving a linear system of equations. We conclude the exposition of our algorithm by showing the following lemma.

**Lemma 2.** *For  $n$  irreducible,  $\alpha^2 + \alpha + 1 \equiv 0 \pmod n$  has solutions if and only if  $d := \deg(n)$  is even.*

( $\Rightarrow$ ) Suppose  $\alpha$  satisfies  $\alpha^2 + \alpha + 1 \equiv 0 \pmod n$ . Then  $1 = \alpha + \alpha^2$ . Squaring each side we get  $1 = \alpha^2 + \alpha^{2^2}$ , then squaring again and again we get  $1 = \alpha^{2^2} + \alpha^{2^3}, \dots$  until  $1 = \alpha^{2^d} + \alpha^{2^{d-1}} = \alpha + \alpha^{2^{d-1}}$ . Summing up these equations we get  $d = 0$ , so  $d$  must be even.

( $\Leftarrow$ ) Now suppose  $d$  is even. Let  $\beta$  be a generator of  $\mathbb{F}_{2^d}^*$  and let  $\alpha = \beta^{\frac{2^d-1}{3}}$ . Then  $\alpha^3 = 1$  and  $\alpha \neq 1$  so  $\alpha^2 + \alpha + 1 = 0$ .  $\square$

*Runtime analysis* We give some estimates for the complexity of our algorithm. Assuming the polynomial  $n$  generated from random  $(b'', m)$  behaves like random polynomials of degree  $k$ , the number of its irreducible factors is asymptotically  $K = \mathcal{O}(\log \deg n)$  [5]. For  $n$  of degree even, we can reasonably approximate the probability that all its factors are of even degree by  $(1/2)^K$ , hence we will need  $2^K = \mathcal{O}(\log n) = \mathcal{O}(\deg p)$  random trials. The factorization of  $n$  can be done in  $\mathcal{O}(\log^{2+\epsilon} n)$  [11] and the continued fraction algorithm is of complexity  $\mathcal{O}(\deg n)$ , so the global complexity of our algorithm is probabilistic polynomial time in  $\deg p$ . Our implementation of the algorithm finds collisions for 1024-bit parameters in a few seconds on a Pentium Intel M processor 1.73GHz.

## 6 Discussion and further work

In this paper, we presented efficient algorithms finding preimages for the LPS hash function and collisions for the Morgenstern hash function with  $q = 2$ . Similar algorithms with the same complexity can be derived for finding preimages for the Morgenstern hash function and for different  $q$  values. Our algorithms build upon the Zémor and Tillich algorithm [15] although they are not trivial extensions of it.

The modified version of LPS hashes proposed by Zémor and Tillich remains unbroken so far regarding both the collision and the preimage properties, as well as their original scheme [12] (with carefully chosen parameters) that used as graph generators  $A_0 = \begin{pmatrix} x & 1 \\ 1 & 0 \end{pmatrix}$  and  $A_1 = \begin{pmatrix} x & x+1 \\ 1 & 1 \end{pmatrix}$ . To avoid our new attack, we suggest modifying the Morgenstern hash function as follows: multiply by  $g_0g_1$  if the bit is 0 and by  $g_0g_2$  if the bit is 1. However, it is not clear what would be the advantages of such a scheme compared to Zémor and Tillich's, as it would not necessarily have better expansion properties, and comparing the graph generators, it will certainly be slower.

In further work, we would like to study the applicability of our algorithm to the Zémor-Tillich (ZT) hash function. The Cayley graphs used in ZT hashes can be naturally embedded into Morgenstern graphs, so our cryptanalysis of Morgenstern hashes might actually open new perspectives on breaking the ZT scheme. Our results may also have applications outside the cryptographic community. The preimage finding algorithm actually solves the diophantine equation (1) which at first sight seems to be a very hard problem. Our path-finding and Zémor and Tillich cycle-finding may improve understanding of LPS graphs when considering their Ihara Zeta-function. Finally, expander graphs have numerous applications in computer science [6], some of which could benefit from our new path-finding algorithm.

Because of all these actual and potential applications, we stress that our algorithms and their running time estimates still may and should be improved in many ways. The algorithm of Section 4 gives paths of length about  $8 \log p$  while the diameter of LPS graphs is known to be  $2 \log p$ . Choosing a smaller  $k$  value in the algorithm will decrease this length and may also improve the running time. Finding other decompositions with less than 4 diagonal matrices is another interesting approach. Finally, adapting our algorithms to make them deterministic is a particularly interesting open problem.

## References

1. [http://csrc.nist.gov/groups/ST/hash/documents/SHA-3\\_FR\\_Notice\\_Nov02\\_2007%20-%20more%20readable%20version.pdf](http://csrc.nist.gov/groups/ST/hash/documents/SHA-3_FR_Notice_Nov02_2007%20-%20more%20readable%20version.pdf).
2. FIPS 180-2 secure hash standard.
3. D. X. Charles, E. Z. Goren, and K. E. Lauter. Cryptographic hash functions from expander graphs. To appear in *Journal of Cryptology*.
4. S. Contini, A. K. Lenstra, and R. Steinfeld. VSH, an efficient and provable collision-resistant hash function. In S. Vaudenay, editor, *EUROCRYPT*, volume 4004 of *Lecture Notes in Computer Science*, pages 165–182. Springer, 2006.
5. P. Flajolet and M. Soria. Gaussian limiting distributions for the number of components in combinatorial structures. *J. Comb. Theory Ser. A*, 53(2):165–182, 1990.
6. S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43:439–561, 2006.
7. A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8:261–277, 1988.

8. V. Lyubashevsky, D. Micciancio, C. Peikert, and A. Rosen. Provably secure FFT hashing. In *NIST 2nd Cryptographic Hash Workshop*, 2006.
9. M. Morgenstern. Existence and explicit construction of  $q + 1$  regular Ramanujan graphs for every prime power  $q$ . *Journal of Combinatorial Theory*, B 62:44–62, 1994.
10. C. Petit, K. Lauter, and J.-J. Quisquater. Full cryptanalysis of lps and morgenstern hash functions. Cryptology ePrint Archive, Report 2008/173, 2008. <http://eprint.iacr.org/>.
11. V. Shoup. On the deterministic complexity of factoring polynomials over finite fields. *Inf. Process. Lett.*, 33(5):261–267, 1990.
12. J.-P. Tillich and G. Zémor. Hashing with  $SL_2$ . In Y. Desmedt, editor, *CRYPTO*, volume 839 of *Lecture Notes in Computer Science*, pages 40–49. Springer, 1994.
13. J.-P. Tillich and G. Zémor. Group-theoretic hash functions. In *Proceedings of the First French-Israeli Workshop on Algebraic Coding*, pages 90–110, London, UK, 1993. Springer-Verlag.
14. X. Wang, Y. L. Yin, and H. Yu. Finding collisions in the full SHA-1. In V. Shoup, editor, *CRYPTO*, volume 3621 of *Lecture Notes in Computer Science*, pages 17–36. Springer, 2005.
15. G. Zémor and J.-P. Tillich. Collisions for the LPS expander graph hash function. To appear in the proceedings of *Advances in Cryptology - EUROCRYPT 2008*, 2008.

## A Toy example of the preimage-finding (path-finding) algorithm in the LPS graph

As an example of our preimage algorithm, we now give a second preimage for the message  $m =$ “This is not for NIST”, when the parameters are  $p = 1125899906842769$  and  $l = 5$ . The ASCII code for  $m$  is 01010100 01101000 01101001 01110011 00100000 01101001 01110011 00100000 01101110 01101111 01110100 00100000 01100110 01101111 01110010 00100000 01001110 01001001 01010011 01010100 which in base 5 gives 3023231443000032312104001244030134 21040324420122212133431310442432021. We start at the identity, with  $g_{IV}$  the identity and  $g_0 = M_1$ . We identify the six graph generators

$$M_{\pm 1} = \begin{pmatrix} 1 \pm 2\mathbf{i} & 0 \\ 0 & 1 \mp 2\mathbf{i} \end{pmatrix}, \quad M_{\pm 2} = \begin{pmatrix} 1 & \pm 2 \\ \mp 2 & 1 \end{pmatrix} \quad M_{\pm 3} = \begin{pmatrix} 1 & 2\mathbf{i} \\ 2\mathbf{i} & 1 \end{pmatrix}$$

with their indices. The function  $\pi$  we choose is given in figure A. The hash value obtained is

$$M = \begin{pmatrix} 1113908155375639 & 815055784352014 \\ 485525153198538 & 30164330826615 \end{pmatrix}.$$

We apply our path-finding algorithm on  $M$ . First, we get a matrix decomposition as in Section 4. After 11 trials, the resulting  $\lambda$ ,  $\alpha$ ,  $\omega$ ,  $\beta_1$  and  $\beta_2$  values

	-3	-2	-1	1	2	3
0	-3	3	2	1	-1	-2
1	2	-3	3	2	1	-1
2	-1	-2	-3	3	2	1
3	1	-1	-2	-3	3	2
4	2	1	-1	-2	-3	3

**Fig. 1.** Table for the function  $\pi$ : the table gives the index of the next matrix for a given current matrix and a given base 5 digit.

are

$$\begin{aligned}\lambda &= 1051846637406052 \\ \alpha &= 698130975272599 \\ \omega &= 846326642296745 \\ \beta_1 &= 150389273084944 \\ \beta_2 &= 480539407839455.\end{aligned}$$

Then we factorize

$$M_\alpha := \begin{pmatrix} 1 & 0 \\ 0 & \alpha \end{pmatrix} = \begin{pmatrix} 349065487636300 + 795285597612250i & 0 \\ 0 & 349065487636300 - 795285597612250i \end{pmatrix}.$$

We choose  $k = 48$ , resulting in  $\lambda = 222458048101540$  and  $m = 11210387681441600668869823936886993015607319565640625$ . After 234 random trials for  $x$ , we finally get  $x = 523712450310834$ ,  $w = 207632734870715$ , and  $n = 4.2489205976128525372183128649803320961$ . The Euclidean algorithm gives us the solution  $y = 2782001231666122912$ ,  $z = 1489057773063985790$ . So the lift of  $M_\alpha$  is

$$M'_\alpha = \left( \begin{array}{c|c} 311426103887630914544037511835 & 3132254927569356406015273012423328 \\ +i766565480745454184887163124346 & +i1676530007976242663293697980252510 \\ \hline -3132254927569356406015273012423328 & 311426103887630914544037511835 \\ +i1676530007976242663293697980252510 & -i766565480745454184887163124346 \end{array} \right).$$

We multiply  $M'_\alpha$  by each of the lifts of the graph generators. Since  $M'_\alpha g'_3$  is divisible by  $l = 5$ ,  $g'_{-3}$  is the last (right-hand) factor of  $M'_\alpha$ . After  $2k$  steps, we get the whole factorization of  $M'_\alpha$ , which we translate into a factorization of  $M_\alpha$  whose indices are 3 -1 2 2 3 1 1 3 1 3 3 3 2 2 3 -1 2 1 1 -3 1 1 1 3 -1 2 -3 2 3 1 -2 -2 -2 1 2 1 1 -3 2 1 2 1 -2 3 -1 3 2 -3 -2 3 1 -2 3 3 2 -3 -1 2 2 2 -1 -3 -1 -3 2 3 1 2 -3 -1 3 2 2 1 3 -2 -3 1 3 -2 -1 -2 3 1 3 2 1 -2 -1 -1 -3 2 1 1 -2 -3. We get the factorizations of  $M_\omega$ ,  $M_{\beta_1}$ , and  $M_{\beta_2}$  the same way. Finally, we put all the pieces of information together and get the sequence -3 -2 1 1 2 -3 -1 -1 -2 1 2 3 1 3 -2 -1 -2 3 1 -3 -2 3 1 2 2 3 -1 -3 2 1 3 2 -3 -1 -3 -1 2 2 2 -1 -3 2 3 3 -2 1 3 -2 -3 2 3 -1 3 -2 1 2 1 2 -3 1 1 2 1 -2 -2 -2 1 3 2 -3 2 -1 3 1 1 1 -3 1 1 2 -1 3 2 2 3 3 3 1 3 1 1 3 2 2 -1 3 2 3 -1 -3 -2 -2 1 3 2 -3 -3 2 3 -2 -3 -3 -2 -1 -2 3 1 1 2 3 2 1 1 -3 1 2

2 1 -2 1 1 2 -3 -2 3 -2 -3 1 1 3 1 2 1 -3 -1 -3 -3 -1 2 3 1 -3 -3 -1 -1 2 -1 3 -2 1 -3  
-3 -1 -2 1 1 -2 -1 -1 3 -2 3 2 2 1 -3 -2 -1 -3 -1 -3 -1 3 2 -1 3 3 -2 -1 -2 1 1 2 2 -3  
-1 3 2 2 -3 -2 -3 -1 -3 1 -3 -2 -1 3 1 -2 3 -2 3 2 1 3 -2 -2 -3 -3 -2 -2 3 -2 -2 -3 -1 3  
1 3 -1 -3 -3 -3 -3 -2 1 3 3 1 -2 3 -1 -2 1 2 -3 1 -2 -2 1 -2 -2 -1 2 2 2 2 2 1 -3 1 1 2  
1 1 3 3 -1 3 3 -2 -1 3 1 2 -1 2 3 -1 2 -3 -2 1 -2 1 1 3 -2 2 -2 -3 -2 -1 3 3 -2 -1 3  
2 -3 2 3 -2 1 1 1 1 -2 1 1 2 -1 3 -1 -2 -1 -2 -1 -2 -2 3 2 1 3 -2 -2 -3 -3 -2 3 3 2 1  
1 1 1 3 2 -1 -3 2 -3 -2 -1 -3 1 -2 -2 -2 1 -2 -1 -2 -1 2 -1 -3 -1 -3 -2 -1 -2 -3 -1 -3  
-1 2 2 3 3 -2 -2 -2 -3 -1 -1 -1 2 -3 -1 3 -1 2 -3 -3 that collides with the original  
message “This is not for NIST”.

## B Collisions for Morgenstern hashes, $q = 2$ and $\deg p(x) = 20$

Now we give a small example for our collision-finding algorithm. The polynomial we choose to target is  $p(x) = x^{20} + x^{17} + x^{14} + x^{13} + x^{12} + x^{11} + x^9 + x^7 + x^5 + x^3 + x^2 + x + 1$ . We choose  $R = 10$  and generate random  $m$  and  $b''$ . After 3 random trials we get  $m = x^9 + x^8 + x^7 + x^6 + x^5 + x^4$ ,  $b'' = x^{10} + x^8 + x^5 + x^2 + 1$  so  $k = 52$ ,  $a = x^{52} + x^{48} + x^{36} + x^{32} + x^{30} + x^{25} + x^{24} + x^{22} + x^{20} + x^{15} + x^{14} + x^{12} + x^{11} + x^{10} + x^9 + x^4 + x^3 + x + 1$ ,  $b = x^{51} + x^{50} + x^{48} + x^{47} + x^{46} + x^{45} + x^{44} + x^{40} + x^{39} + x^{38} + x^{37} + x^{36} + x^{35} + x^{34} + x^{32} + x^{31} + x^{30} + x^{29} + x^{28} + x^{27} + x^{25} + x^{24} + x^{23} + x^{22} + x^{20} + x^{19} + x^{18} + x^{16} + x^{11} + x^{10} + x^9 + x^8 + x^5 + x^4 + x^3 + x^2 + x$  and  $n = x^{62} + x^{61} + x^{59} + x^{57} + x^{55} + x^{53} + x^{52} + x^{51} + x^{50} + x^{49} + x^{48} + x^{46} + x^{45} + x^{40} + x^{37} + x^{31} + x^{29} + x^{28} + x^{26} + x^{25} + x^{24} + x^{23} + x^{16} + x^{15} + x^{13} + x^{12} + x^{10} + x^6 + x^5 + x^3 + 1$ .

The polynomial  $n$  has three factors  $n_1 = x^{56} + x^{54} + x^{53} + x^{50} + x^{48} + x^{46} + x^{44} + x^{40} + x^{36} + x^{34} + x^{33} + x^{30} + x^{29} + x^{22} + x^{20} + x^{18} + x^{13} + x^{11} + x^7 + x^6 + x^5 + x^3 + 1$ ,  $n_2 = x^4 + x^3 + x^2 + x + 1$  and  $n_3 = x^2 + x + 1$  which are all of even degrees. For each factor  $n_i$  we compute  $\alpha$  such that  $\alpha^2 + \alpha + 1 \equiv 0 \pmod{n_i}$  and use this value and the continued fraction algorithm to recover  $(c_i, d_i)$  such that  $c_i^2 + d_i^2 + c_i d_i \equiv 0 \pmod{n_i}$ : we get  $(c_1, d_1) = (x^{26} + x^{25} + x^{24} + x^{21} + x^{20} + x^{18} + x^{16} + x^{14} + x^{13} + x^{11} + x^8 + x^6 + x^5 + x + 1, x^{28} + x^{23} + x^{21} + x^{19} + x^{15} + x^{13} + x^{10} + x^7 + x^5 + x^4 + x^2 + x + 1)$ ,  $(c_2, d_2) = (x, x^2 + 1)$  and  $(c_3, d_3) = (x, 1)$ .

Combining these partial results we get  $c = x^{51} + x^{50} + x^{47} + x^{41} + x^{40} + x^{36} + x^{31} + x^{27} + x^{26} + x^{25} + x^{24} + x^{23} + x^{22} + x^{20} + x^{18} + x^{17} + x^{16} + x^{14} + x^{12} + x^{11} + x^{10} + x^9 + x^7 + x^4$  and  $d = x^{51} + x^{50} + x^{49} + x^{48} + x^{47} + x^{45} + x^{44} + x^{43} + x^{42} + x^{39} + x^{36} + x^{33} + x^{31} + x^{30} + x^{29} + x^{27} + x^{26} + x^{25} + x^{22} + x^{21} + x^{20} + x^{18} + x^{17} + x^{16} + x^{14} + x^{13} + x^9 + x^7 + 1$ .

We can verify that

$$(a^2 + b^2 + ab) + (c^2 + d^2 + cd)x = (1 + x)^{2k}$$

and  $(a, b, c, d) \equiv (1 + x)^k(1, 0, 0, 0) \pmod{p}$ . We factorize the lifted matrix and, using the indices of the generators given in Section 5, we get the following collision with the void message: 0 2 0 2 0 1 2 1 2 1 2 1 2 0 2 0 2 0 2 0 2 0 1 0 2 0 2  
1 0 2 1 0 1 0 2 1 0 1 2 1 2 1 2 1 0 2 1 0 1 2 0 1 0 1 0 1 0 2 1 0 1 2 0 2 1 2 0 2 0 1  
2 0 2 0 1 0 2 1 2 1 0 2 0 1 0 1 2 0 2 0 2 0 2 0 1 2 1 0 2 0 2 1 0 1.