Algebraic Cryptanalysis of a Quantum Money Scheme The Noise-Free Case

Marta Conde Pena, Jean-Charles Faugère, Ludovic Perret

Algebraic Cryptanalysis of a Quantum Money Scheme The Noise-Free Case

Marta Conde Pena, Jean-Charles Faugère[2,1,3], and Ludovic Perret[1,2,3]

[1] Institute of Physical and Information Technologies (ITEFI) – Spanish National Research Council (CSIC)
[2] Sorbonne Universités, UPMC Univ Paris 06, POLSYS, UMR 7606, LIP6, F-75005, Paris, France
[3] INRIA, Paris-Rocquencourt Center, POLSYS Project
[4] CNRS, UMR 7606, LIP6, F-75005, Paris, France
marta.conde@iec.csic.es,jean-charles.faugere@inria.fr,ludovic.perret@lip6.fr

Abstract. We investigate the Hidden Subspace Problem (HSP_q) over $\mathbb{F}_q$:
Input : $p_1, \ldots, p_m, q_1, \ldots, q_m \in \mathbb{F}_q[x_1, \ldots, x_n]$ of degree $d \geq 3$ (and $n \leq m \leq 2n$).
Find : a subspace $A \subset {\mathbb{F}_q}^n$ of dimension $n/2$ (n is even) such that

$$p_i(A) = 0 \ \forall i \in \{1, \ldots, m\} \text{ and } q_j(A^{\perp}) = 0 \ \forall j \in \{1, \ldots, m\},$$

where $A^{\perp}$ denotes the orthogonal complement of A with respect to the usual scalar product in $\mathbb{F}_q$.
This problem underlies the security of the first public-key quantum money scheme that is proved to be cryptographically secure under a non quantum but classic hardness assumption. This scheme was proposed by S. Aaronson and P. Christiano [1] at STOC'12. In particular, it depends upon the hardness of HSP_2. More generally, Aaronson and Christiano left as an open problem to study the security of the scheme for a general field $\mathbb{F}_q$. We present a randomized polynomial-time algorithm that solves the HSP_q for $q > 2$ with success probability $\approx 1 - 1/q$. So, the quantum money scheme extended to $\mathbb{F}_q$ is not secure. Finally, based on experimental results and a structural property of the polynomials that we prove, we conjecture that there is also a randomized polynomial-time algorithm solving the HSP_2 with high probability. To support our theoretical results, we also present several experimental results confirming that our algorithms are very efficient in practice. We emphasize that [1] proposes a non-noisy and a noisy version of the public-key quantum money scheme. The noisy version of the quantum money scheme remains secure.

1 Introduction

The no-cloning theorem in quantum mechanics states the impossibility of creating identical copies of an unknown arbitrary quantum money state. In [20], Wiesner suggested to take advantage of this physical law in order to construct a scheme for (quantum) money that could not be counterfeited. The initial work of Wiesner has been then followed by several papers that try to improve the initial idea of [20], i.e. [5,17,18]. This line of research culminated with the proposal of Aaronson and Christiano [1] at STOC'12 who proposed a public-key quantum money scheme.

A public-key quantum money scheme is a scheme in which anyone with a quantum device can verify if a banknote is valid rather than only the bank that issued (in contrast to [20]). A public-key quantum money scheme based on knot theory was introduced in [12]. However, its security is not well understood. The scheme proposed by Aaronson and Christiano in [1] is the first that is public-key and proved to be cryptographically secure under a classical (as in non-quantum) hardness assumption. The scheme is based on hiding two orthogonal subspaces by expressing each of them as the common zeros of a set of appropriate random multivariate non-linear polynomials. In particular, its security relies on the assumption that the following problem is hard:

Hidden Subspaces Problem (HSP_q)
Input : polynomials $p_1, \ldots, p_m, q_1, \ldots, q_m \in \mathbb{F}_q[x_1, \ldots, x_n]$ of degree $d \geq 3$, $n \leq m \leq 2n$.
Find : a subspace $A \subset {\mathbb{F}_q}^n$ of dimension $n/2$ (n is even) such that

$$p_i(A) = 0 \ \forall i \in \{1, \ldots, m\} \text{ and } q_j(A^{\perp}) = 0 \ \forall j \in \{1, \ldots, m\},$$

where $A^{\perp}$ denotes the orthogonal complement of A with respect to the standard scalar product in $\mathbb{F}_q$.

We emphasize that in [1], the authors propose a non-noisy and a noisy version of the public-key quantum money scheme. In this paper we only consider the noise-free version of the quantum money scheme.

In particular, the non-noisy version of the quantum money scheme relies on the HSP_2, and Aaronson and Christiano conjecture that it cannot be solved in polynomial-time. They also state as an open problem the study of the scheme extended to a general field $\mathbb{F}_q$, which brings up the question of the hardness of HSP_q.

We analyze the hardness of the HSP_q. The main idea is to model the problem as a set of algebraic equations. Expressing elements as the common zeroes of a set of random multivariate non-linear polynomials is the core of algebraic attacks, e.g. [15,11,16]. However, in this case we can exploit that there are two sets of public polynomials whose sets of zeros are two subspaces orthogonal to each other.

Aside from this quantum money scheme, the HSP_q has also interest as a general computer algebra problem closely related to the isomorphism of polynomials [19]. Given $\mathbf{p} = T \circ \mathbf{p}' \circ S$, where $\mathbf{p} = (p_1, \ldots, p_m), \mathbf{p}' = (p'_1, \ldots, p'_m) \in \mathbb{F}_q[x_1, \ldots, x_n]^m$ and T, S are affine invertible transformations, the *Isomorphism of Polynomials* (IP) problem consists on recovering T and S. The HSP_q can be seen as a slight modification of the isomorphism of polynomials problem where $\mathbf{p} = \mathbf{0}$, T is the identity transformation and S is linear but not invertible.

1.1 Main Results

Our results mostly rely on Gröbner bases and linear algebra techniques. This is because we are capable of identifying the solution of the HSP_q as the unique solution of an overdetermined system of multivariate equations in $N = n^2/4$ unknowns (Section 3.1, Proposition 5). The properties of this system are different for $q = 2$ and $q > 2$, so we study separately both cases.

Our first main result (Section 3) solves an open problem presented in [1], which is the study of the HSP_q for $q > 2$. From the algebraic equations describing HSP_q, we observe that we can extract a set of linear equations (Lemma 3). Due to the shape of the linear equations, we can prove prove that sufficiently many linearly independent ones can be extracted. This gives:

Theorem 1 (Section 3.2). *Let $N = n^2/4$. There is a randomized polynomial-time algorithm solving HSP_q, for $q > 2$, with complexity $\mathcal{O}(N^\omega) = \mathcal{O}(n^{2\omega})$, where $2 \leq \omega \leq 3$ is the linear algebra constant, and success probability*

$$\frac{\gamma_q(n/2)\gamma_q(m)}{\gamma_q(m - n/2)},$$

$\gamma_q(k)$ being the probability that a random $k \times k$ matrix with entries in $\mathbb{F}_q$ is invertible. For n big enough, the success probability is $\approx 1 - 1/q$.

In Section 3.3, we report experimental results demonstrating that HSP_q, with $q > 2$, can be solved very efficiently. For $n \leq 20$, the algorithm requires less than 0.1 s. for various q. We have implemented the algorithm using the Magma software [6]. The code is provided with the submission so that the results can be reproduced or conducted for bigger values of n.

Our second result is concerned with the HSP_2 (Section 4). In this case our system does not contain, except with a small probability, linear equations and so the approach needs to be different. Still, in the case of the HSP_2, we have an algebraic system of equations which is very overdetermined.

Proposition 1 (Section 4). *Let $\big(\mathbf{p} = (p_1, \ldots, p_m), \mathbf{q} = (q_1, \ldots, q_m)\big) \in \mathbb{F}_2[\boldsymbol{x}]^m \times \mathbb{F}_2[\boldsymbol{x}]^m$ be degree-d multivariate polynomials. Let $A \subset \mathbb{F}_2^n$ be a vector subspace of dimension $n/2$. If A is a solution of HSP_2 on $(\mathbf{p}, \mathbf{q}) \in \mathbb{F}_2[\boldsymbol{x}]^m \times \mathbb{F}_2[\boldsymbol{x}]^m$, then we can construct an algebraic system of equations $\mathrm{Sys}_{\mathrm{HSP}_2}$ over $\mathbb{F}_2$ in $N = n^2/4$ variables with at most $2m\left[\binom{n/2}{1} + \binom{n/2}{2} + \ldots + \binom{n/2}{d}\right]$ such that a systematic basis of A*

vanishes $\mathrm{Sys}_{\mathrm{HSP}_2}$ *with probability* $\gamma_2(n/2)$*, where* $\gamma_2(n/2)$ *denotes the probability of a random matrix with entries in* $\mathbb{F}_2$ *of being invertible. For* n *big enough, this is* $\approx 1/2$*.*

So, we can still hope that the computation of a Gröbner basis of the system will be efficient. In Section 4.1, we run experiments to confirm this intuition. It appears that $\mathrm{Sys}_{\mathrm{HSP}_2}$ is much easier to solve than a semi-regular system of the same size. The MAGMA code of this part is also provided with the submission. Typically, we can solve in practice $\mathrm{Sys}_{\mathrm{HSP}_2}$ for $n \leq 18$ and $d = 3$ in less than 3 hours (for smaller n, we can solve in few minutes). For $n = 18$, we have to solve a system of degree-3 equations with 81 variables. In practice, we observed that the maximum degree reached during the computation of a Gröbner basis of $\mathrm{Sys}_{\mathrm{HSP}_2}$ is bounded from above by a small constant. Based on this observation we conjecture then that:

Conjecture 1. The degree of regularity is bounded above by $d + 1$.

If this conjecture is true, the following result is obtained:

Theorem 2. *Let* $N = n^2/4$*. There is a randomized polynomial-time algorithm solving* HSP_2 *with a complexity of* $\mathcal{O}(N^{\omega(d+1)}) = \mathcal{O}(n^{2\omega(d+1)})$*, where* $2 \leq \omega \leq 3$ *is the linear algebra constant, and success probability* $\gamma_2(n/2)$*, where* $\gamma_2(k)$ *denotes the probability of a random* $k \times k$ *matrix with entries in* $\mathbb{F}_2$ *of being invertible. For* n *big enough, the success probability of the algorithm is* $\approx 1/2$*.*

To support our assumption we analyze in Section 5 the structure of $\mathrm{Sys}_{\mathrm{HSP}_2}$. We prove a structural property (due to the orthogonality of the hidden subspaces) that allows to obtain equations of degree lower than d from the public polynomials $\mathrm{Sys}_{\mathrm{HSP}_2}$ by performing simple manipulations on the initial system. In particular:

Proposition 2. *We can easily generate* $\mathcal{O}(m^2)$ *equations of degree* $d-1$ *which are linear combinations of the equations from* $\mathrm{Sys}_{\mathrm{HSP}_2}$*.*

This means that a Gröbner computation on $\mathrm{Sys}_{\mathrm{HSP}_2}$ will generate at the very first step many equations of lower degree. This is known as a fall of degree and it is typically a behaviour which is not occurring in a random (i.e. semi-regular) system of equations. So, it is a first step towards proving our conjecture.

1.2 Organization of the Paper

In Section 2 we introduce some notation that will be used throughout the paper, we recall the basics of Gröbner bases and we describe the non-noisy version of the quantum money scheme of [1]. The first part of Section 3 is concerned with the general modeling of the HSP_q as a system of multivariate non-linear equations, and the second part is dedicated to obtain the algorithm of the first Theorem. Sections 4 and 5 are dedicated to the HSP_2. In Section 4 we explain precisely why the behaviour of HSP_q is different for $q = 2$ and $q > 2$, we report experimental results and we derive our conjecture which, if true, results in the second Theorem. Section 5 is the most technical one, in which we explain that equations of degree $< d$ can be obtained due to the orthogonality of the hidden subspaces.

2 Preliminaries

We first recall some basics of Gröbner bases as the main tool to approach non-linear systems. Then we describe precisely our target problem and its relation with the quantum money scheme proposed in [1]. Before that we fix some general notation: we denote by $\mathbb{F}_q$ the finite field with q elements, we set $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbb{F}_q[\mathbf{x}] = \mathbb{F}_q[x_1, \ldots, x_n]$ to be the polynomial ring over $\mathbb{F}_q$ in the unknowns $x_1, \ldots, x_n$. We denote by $\mathrm{M}(\mathbb{F}_q[\mathbf{x}])$ the set of monomials in $\mathbb{F}_q[\mathbf{x}]$, $\mathrm{M}(\mathbb{F}_2[\mathbf{x}])$ refers to the set of square-free monomials in $\mathbb{F}_2[\mathbf{x}]$ and $\mathrm{M}_s(\mathbb{F}_q[\mathbf{x}])$ refers to the set of monomials of degree s in $\mathbb{F}_q[\mathbf{x}]$. As usual, $\mathcal{M}_{k,\ell}(\mathbb{F}_q)$ denotes the set of $k \times \ell$ matrices with entries in $\mathbb{F}_q$, $\mathcal{M}_k(\mathbb{F}_q)$ denotes the square matrices of order k with entries in $\mathbb{F}_q$, and $\mathrm{GL}_k(\mathbb{F}_q)$ the set of invertible matrices in $\mathcal{M}_k(\mathbb{F}_q)$.

2.1 Basics of Computer Algebra

As systems of multivariate non-linear equations are the key component of this work we recall some aspects of Gröbner bases computations [10,9,8]. Given a polynomial ideal over $\mathbb{F}_q$, say $\mathcal{I} = \langle f_1, \ldots, f_s \rangle = \{\sum_{i=1}^{s} f_i h_i \mid h_1, \ldots, h_s \in \mathbb{F}_q[\mathbf{x}]\}$, Gröbner bases provide a way to obtain the variety $V_q(\mathcal{I}) = \{x \in \mathbb{F}_q \mid f_i(\mathbf{x}) = 0, \text{for all } 1 \leq i \leq s\}$ by transforming the initial generators of the ideal into new generators with better properties, in the sense that computing the variety becomes simpler (this "better" set of generators is precisely the Gröbner basis).

The classic method to compute Gröbner bases is Buchberger's algorithm [10,9,8], but more efficient methods, such as F_4 [13] and F_5 [14], have been proposed. F_5 is considered to be one of the most efficient algorithms up to date for computing Gröbner bases. It uses linear algebra techniques and suppresses useless computations carried out in Buchberger's algorithm.

For increasing values of $\tilde{d}$, the F_5 algorithm successively reduces to row echelon form matrices of the form

$$A_{\tilde{d}} = \begin{array}{c} m_1 f_1 \\ m_2 f_2 \\ m_3 f_3 \\ \ldots \end{array} \begin{pmatrix} \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{pmatrix} \tag{1}$$

where the columns are indexed by the monomials ordered decreasingly with respect to $<$, and m_i are monomials such that $\deg(m_j f_i) \leq \tilde{d}$. At some point, for some d, the reduced row echelon form of the matrix A_d contains a Gröbner basis. This maximum degree d reached during a F_5 computation is called the *degree of regularity* and it is an important parameter when assessing the running time of a Gröbner basis computation.

Systems verifying certain hypotheses are called semi-regular [2,4], and they are interesting due to two reasons. Firstly, because if a system is chosen at random it turns out to be semi-regular with high probability, and secondly because the degree of regularity is known for this kind of systems. In fact, if $\{f_1, \ldots, f_s\} \subset \mathbb{F}_q[\mathbf{x}]$ is a semi-regular system, where each f_i has degree d_i, its degree of regularity is given by the index of the first non-positive coefficient of the power series

$$\sum_{k \geq 0} c_k z^k = \frac{\prod_{i=1}^{s} (1 - z^{d_i})}{(1-z)^n}. \tag{2}$$

The time complexity of computing a Gröbner basis [3] is roughly given by the time spent carrying out the row echelon reduction of A_d, which is $\mathcal{O}((\#A_d)^\omega)$, where $2 \leq \omega \leq 3$ is the linear algebra constant. Since the size of the matrix A_d can be roughly approximated by $\mathcal{O}\left(n^d\right)$, this gives an overall complexity of $\mathcal{O}\left(n^{\omega d}\right)$.

2.2 Definition of the Problem

From now on we will always assume that n is even. Recall that we are focusing on analyzing the hardness of the following problem:

Hidden Subspaces Problem (HSP_q)
Input : $p_1, \ldots, p_m, q_1, \ldots, q_m \in \mathbb{F}_q[\mathbf{x}]$ of degree $d \geq 3$ ($n \leq m \leq 2m$).
Find : a subspace $A \subset {\mathbb{F}_q}^n$ of dimension $n/2$ such that

$$p_i(A) = 0 \ \forall i \in \{1, \ldots, m\} \text{ and } q_j(A^\perp) = 0 \ \forall j \in \{1, \ldots, m\},$$

where $A^{\perp}$ denotes the orthogonal complement of A with respect to the standard scalar product in $\mathbb{F}_q$.

As mentioned in the introduction, HSP_q arises in relation to the security of the non-noisy version of the quantum money scheme proposed in [1]. The private key of this scheme is a subspace $A \subset {\mathbb{F}_q}^n$, and the polynomials $p_1, \ldots, p_m \in \mathbb{F}_q[\mathbf{x}]$ (vanishing on A) and $q_1, \ldots, q_m \in \mathbb{F}_q[\mathbf{x}]$ (vanishing on $A^{\perp}$) are the public key. To output money, the bank queries an oracle to obtain a basis of A and using this description of the subspace it generates a quantum state \$ which is the banknote. The verifying process is based on the fact that it is easy to check whether a given element is a zero of a polynomial or not.

The recovery of A compromises the security of the scheme, so it becomes crucial that the HSP_q cannot be easily solved. It is conjectured in [1] that, for big enough d, there is no polynomial-time algorithm that solves HSP_2 with success probability $\Omega\left(2^{-n/2}\right)$.

Before proceeding any further we need to detail how the keys are generated. This is specified in [1]. The generation of an uniformly random subspace is clear by just choosing a full rank matrix in $\mathcal{M}_{n/2,n}(\mathbb{F}_q)$. The generation of an uniformly random polynomial vanishing on a given subspace can be done in $\mathcal{O}\left(n^d\right)$-time

Lemma 1 ([1]). *Denote by $I_{d,A}$ the set of polynomials of degree d that vanish on A, by $e_i \in {\mathbb{F}_q}^n$ the vector that has a 1 in its i-th position and 0 elsewhere, and by E the subspace generated by the vectors $e_1, \ldots, e_{n/2}$. We have:*

1. *A polynomial is in $I_{d,E}$ if and only if each of its monomials is divisible by an element in the set $\{x_{n/2+1}, \ldots, x_n\}$.*
2. *If L is an invertible linear transformation on $I_{d,A}$, the function $p(\boldsymbol{x}) \to p(\boldsymbol{x}L)$ maps $I_{d,A}$ to $I_{d,AL^{-1}}$.*

Applying this lemma, one can generate polynomials vanishing on the appropriate subspace in the following way:

Proposition 3. ***(Vanishing polynomial)*** *The generation of an uniformly random polynomial of degree d vanishing on a given subspace A consists of the following two steps:*

1. *Generate a polynomial $p(\boldsymbol{x})$ of degree d vanishing on E: by lemma 1(1), this is done including each monomial of degree d or lower independently and with probability $1/2$ if it is divisible by an element in the set $\{x_{n/2+1}, \ldots, x_n\}$.*
2. *Transform the polynomial $p(\boldsymbol{x})$ into one vanishing on A: considering the matrix L of change of basis (i.e., $E = AL$), the polynomial $p(\boldsymbol{x}L)$ vanishes on A by lemma 1(2).*

We have preformed all our experiments using Proposition 3.

3 The HSP_q, for $q > 2$

We analyze the hardness of the HSP_q for $q > 2$. This is an open problem in [1] that arises when studying the security of the quantum money scheme extended to a general field $\mathbb{F}_q$. We conclude that the quantum money scheme extended to $\mathbb{F}_q$ is not secure. First we show that, with a certain probability, the HSP_q can be modeled by a suitable set of non-linear equations. Then we prove that, with very high probability, enough linear equations that are linearly independent can be extracted from it. This results in a randomized polynomial-time algorithm for the HSP_q ($q > 2$).

3.1 General Modeling of HSP_q

In this part we show that the HSP_q can be rather naturally modeled as a set of algebraic equations. The first straightforward modeling presented is however not optimal as it includes many equivalent solutions. We show how we can use the structure of our problem to remove the unnecessary solutions.

We abuse notation and denote by A either a subspace of $\mathbb{F}_q^n$ of dimension $n/2$ or a matrix in $\mathcal{M}_{n/2,n}(\mathbb{F}_q)$ whose rows are the elements of a basis of the subspace A.

Proposition 4. *Let* $\big(\mathbf{p} = (p_1, \ldots, p_m), \mathbf{q} = (q_1, \ldots, q_m)\big) \in \mathbb{F}_q[\boldsymbol{x}]^m \times \mathbb{F}_q[\boldsymbol{x}]^m$ *be a degree-d instance of* HSP_q*. Let* $\big(y_1, \ldots, y_{n/2}\big)$ *be variables and* G *(resp.* $G^\perp$*) be a formal matrix of size* $n/2 \times n$ *(resp.* $n/2 \times n$*). We consider the system:*

$$\mathrm{SysNaive}_{\mathrm{HSP}_q} = \big\{\mathrm{Coeff}(p_i, t), \mathrm{Coeff}(q_j, t) \mid \forall i, j \in \{1, \ldots, m\}, \forall t \in M\big(\mathbb{F}_q[y_1, \ldots, y_{n/2}]\big)\big\} \tag{3}$$

where $\mathrm{Coeff}(p_i, t)$ *denotes the coefficient of* $t \in \mathrm{M}(\mathbb{F}_q[g_1, \ldots, g_N])$ *in* $p_i\big((y_1, \ldots, y_{n/2}) \cdot G\big)$ *and* $\mathrm{Coeff}(q_j, t)$ *the coefficient of* $t \in \mathrm{M}(\mathbb{F}_q[g_1^\perp, \ldots, g_N^\perp])$ *in* $q_j\big((y_1, \ldots, y_{n/2}) \cdot G^\perp\big)$*.* $\mathrm{SysNaive}_{\mathrm{HSP}_q}$ *is a system of* $\mathcal{O}(n^d)$ *algebraic equations over* $\mathbb{F}_q$ *in* n^2 *variables (the entries of* G *and* $G^\perp$*).*

Let $A \subset \mathbb{F}_q^n$ *be a vector subspace of dimension* $n/2$*. If* A *is a solution of* HSP_q *on* $(\mathbf{p}, \mathbf{q}) \in \mathbb{F}_q[\boldsymbol{x}]^m \times \mathbb{F}_q[\boldsymbol{x}]^m$ *then the components of* A *and* $A^\perp$ *vanish all the equations* $\mathrm{SysNaive}_{\mathrm{HSP}_q}$*.*

Proof. This is an immediate consequence of the fact that every element of the subspace A (resp. $A^\perp$) can be expressed as $(y_1, \ldots, y_{n/2})A$ (resp. $(y_1, \ldots, y_{n/2})A^\perp$). As a consequence, all the coefficients of the polynomials $p_i\big((y_1, \ldots, y_{n/2}) \cdot A\big)$ (resp. $p_i\big((y_1, \ldots, y_{n/2}) \cdot A^\perp\big)$ must be equal to zero. □

It is easy to see that $\mathrm{SysNaive}_{\mathrm{HSP}_q}$ has many solutions which are equivalent. If a vector subspace $A \subset \mathbb{F}_q^n$ is a solution of HSP_q, then any basis of A will be a solution $\mathrm{SysNaive}_{\mathrm{HSP}_q}$. It is then natural to define a canonical form of the solutions of HSP_q.

Lemma 2. *Let* $\big(\mathbf{p} = (p_1, \ldots, p_m), \mathbf{q} = (q_1, \ldots, q_m)\big) \in \mathbb{F}_q[\boldsymbol{x}]^m \times \mathbb{F}_q[\boldsymbol{x}]^m$ *be a degree-d instance of* HSP_q*. Let* $A \subset \mathbb{F}_q^n$ *be a vector subspace of dimension* $n/2$*. If* A *is a solution of* HSP_q *on* $(\mathbf{p}, \mathbf{q}) \in \mathbb{F}_q[\boldsymbol{x}]^m \times \mathbb{F}_q[\boldsymbol{x}]^m$*, then for any* $S \in \mathrm{GL}_{n/2}(\mathbb{F}_q)$*,* $S \cdot A$ *is a solution of* HSP_q *on* $(\mathbf{p}, \mathbf{q}) \in \mathbb{F}_q[\boldsymbol{x}]^m \times \mathbb{F}_q[\boldsymbol{x}]^m$*.*

Proof. For all $i, 1 \leq i \leq m$, $p_i\big((y_1, \ldots, y_{n/2})S \cdot A\big) = 0$ holds as a consequence of S being invertible. Also, since $(SA)^\perp = A^\perp$ (as $A^\perp (SA)^T = A^\perp A^T S^T = 0$ considering that $A^\perp A^T = 0$), it also holds that $q_j\big((y_1, \ldots, y_{n/2})(SA)^\perp\big) = 0$ for all $j, 1 \leq j \leq m$. □

A direct consequence of Lemma 2 is that we can assume, with high probability, that a vector space solution A of HSP_q is given in systematic form. This is, we can suppose that $A = (I|G)$, where $G \in \mathrm{GL}_{n/2}(\mathbb{F}_q)$ and I is the $n/2 \times n/2$ identity matrix. If A has such a form then $A^\perp = (G^T|I)$.

Fact 1 *We recall that the probability that a random matrix in* $\mathcal{M}_n(\mathbb{F}_q)$ *is invertible is given by:*

$$\gamma_q(n) = \prod_{i=1}^{n} \left(1 - \frac{1}{q^i}\right).$$

It is well known that:

$$\lim_{n \to \infty} \gamma_q(n) = 1 - \frac{1}{q} + \mathcal{O}\left(\frac{1}{q^2}\right).$$

For big values of q*,* $\gamma_q(n/2)$ *is close to* 1*, which justifies the restriction on the shape of the subspace* A*.*

We can now improve the modeling thanks to a canonical form of the solutions. We remove all the solutions of $\mathrm{SysNaive}_{\mathrm{HSP}_q}$ which correspond to equivalent bases. To do so, we generate a similar system of equations but with a smaller number of variables.

Proposition 5. *Let* $\big(\mathbf{p} = (p_1, \ldots, p_m), \mathbf{q} = (q_1, \ldots, q_m)\big) \in \mathbb{F}_q[\boldsymbol{x}]^m \times \mathbb{F}_q[\boldsymbol{x}]^m$ *be a degree-d instance of* HSP_q*. Let* $\big(y_1, \ldots, y_{n/2}\big)$ *be variables and* G *be a formal matrix of size* $n/2 \times n/2$ *and* $N = n^2/4$*. We consider the system:*

$$\mathrm{Sys}_{\mathrm{HSP}_q} = \big\{\mathrm{Coeff}(p_i, t), \mathrm{Coeff}(q_j, t) \mid \forall i, j \in \{1, \ldots, m\}, \forall t \in M\big(\mathbb{F}_q[y_1, \ldots, y_{n/2}]\big)\big\} \tag{4}$$

where $\mathrm{Coeff}(p_i, t)$ *denotes the coefficient of* $t \in \mathrm{M}(\mathbb{F}_q[g_1, \ldots, g_N])$ *in* $p_i\big((y_1, \ldots, y_{n/2}) \cdot (I|G)\big)$*, and* $\mathrm{Coeff}(q_j, t)$ *the coefficient of* $t \in \mathrm{M}(\mathbb{F}_q[g_1, \ldots, g_N])$ *in* $q_j\big((y_1, \ldots, y_{n/2}) \cdot (G^T|I)\big)$*.* $\mathrm{Sys}_{\mathrm{HSP}_q}$ *is a system of* $\mathcal{O}(n^d)$ *algebraic equations over* $\mathbb{F}_q$ *in* N *variables (the entries of* G*).*

Let $A \subset \mathbb{F}_q^n$ *be a vector subspace of dimension* $n/2$*. If* A *is a solution of* HSP_q *on* $(\mathbf{p}, \mathbf{q}) \in \mathbb{F}_q[\boldsymbol{x}]^m \times \mathbb{F}_q[\boldsymbol{x}]^m$*, then* A *admits with probability* $\gamma_q(n/2)$ *a basis in systematic form whose components vanish all the equations* $\mathrm{Sys}_{\mathrm{HSP}_q}$*.*

Proof. This follows easily from Lemma 2 and Proposition 4. □

So, Lemma 2 permitted to divide by 4 the number of variables that we have to consider.

3.2 Randomized Polynomial-Time Algorithm for HSP_q, with $q > 2$

According to Proposition 5, solving HSP_q is equivalent with high probability (w.h.p) to solve the non-linear system $\mathrm{Sys}_{\mathrm{HSP}_q}$. In this part we show that the non-linear system can be solved in polynomial-time. This is due to the fact that we can extract from $\mathrm{Sys}_{\mathrm{HSP}_q}$ sufficiently many linear equations that are linearly independent.

Lemma 3. *Let* $\big(\mathbf{p} = (p_1, \ldots, p_m), \mathbf{q} = (q_1, \ldots, q_m)\big) \in \mathbb{F}_q[\boldsymbol{x}]^m \times \mathbb{F}_q[\boldsymbol{x}]^m$ *be a degree-d instance of* HSP_q*. Let* $p_i^{(1)}$ *(resp.* $q_i^{(1)}$*) be the homogeneous component of degree* 1 *of* p_i *(resp.* q_i*), that is:*

$$p_i{}^{(1)} = \sum_{j=1}^{n} \lambda_{i,j}^p x_j, \quad \textit{where } \lambda_{i,1}, \ldots, \lambda_{i,n} \in \mathbb{F}_q,$$

$$q_i{}^{(1)} = \sum_{j=1}^{n} \lambda_{i,j}^q x_j, \quad \textit{with } \lambda_{i,1}, \ldots, \lambda_{i,n} \in \mathbb{F}_q.$$

For $i \in \{1, \ldots, m\}$ *and* $k \in \{1, \ldots, n/2\}$*, the linear equations:*

$$\begin{cases} \sum_{j=1}^{n/2} \lambda_{i,j+n/2}^p g_{j+n/2(k-1)} + \lambda_{i,k}^p \\ \sum_{j=1}^{n/2} \lambda_{i,j}^q g_{k+n/2(j-1)} + \lambda_{i,k+n/2}^q \end{cases}$$

are in $\mathrm{Sys}_{\mathrm{HSP}_q}$*.*

Proof. Let G be a formal matrix of size $n/2 \times n/2$. If we expand the products $(y_1, \ldots, y_{n/2})(I|G)$ and $(y_1, \ldots, y_{n/2})(G^T|I)$, we can see that $\mathrm{Sys}_{\mathrm{HSP}_q}$ is obtained from the coefficients of:

$$\begin{cases} p_i\left(y_1, \ldots, y_{n/2}, \sum_{t=0}^{n/2-1} g_{tn/2+1} y_{t+1}, \ldots, \sum_{t=0}^{n/2-1} g_{tn/2+n/2} y_{t+1}\right), \forall i, 1 \le i \le m, \\ q_j\left(\sum_{t=1}^{n/2} g_t y_t, \ldots, \sum_{t=1}^{n/2} g_{n/2(n/2-1)+t} y_t, y_1, \ldots, y_{n/2}\right), \forall j, 1 \le j \le m. \end{cases} \tag{5}$$

It is clear that the equations $\mathrm{Coeff}(p_i, y_1), \ldots, \mathrm{Coeff}(p_i, y_{n/2})$ on the one hand, and the equations $\mathrm{Coeff}(q_i, y_1), \ldots, \mathrm{Coeff}(q_i, y_{n/2})$ on the other hand, are linear for all $i \in \{1, \ldots, m\}$. Taking into account the expressions of $p_i^{(1)}$ and $q_i^{(1)}$ as well as (5), we have that for $k = 1, 2, \ldots, n/2$:

$$\mathrm{Coeff}(p_i, y_k) = \lambda_{i,k}^p + \sum_{j=1}^{n/2} \lambda_{i,j+n/2}^p g_{j+(k-1)n/2},$$

$$\mathrm{Coeff}(q_i, y_k) = \lambda_{i,k+n/2}^q + \sum_{j=1}^{n/2} \lambda_{i,j}^q g_{k+(j-1)n/2},$$

as required. □

Let $N = n^2/4$. Since $m \geq n$, the system of linear equations in Lemma 3 is already overdetermined with at most $2mn/2 = mn \geq 4N$ linear equations versus N unknowns. We show now that among these (at most) mn linear equations there are, with high probability, at least N linearly independent ones, enough to solve it.

Lemma 4. *Let* $\big(\mathbf{p} = (p_1, \ldots, p_m), \mathbf{q} = (q_1, \ldots, q_m)\big) \in \mathbb{F}_q[\boldsymbol{x}]^m \times \mathbb{F}_q[\boldsymbol{x}]^m$ *be a degree-d instance of* HSP_q. *With probability* $\frac{\gamma_q(m)}{\gamma_q(m-n/2)}$, *we can extract from* $\mathrm{Sys}_{\mathrm{HSP}_q}$ *at least* $N = n^2/4$ *linear equations that are linearly independent.*

Proof. The $mn \times N$ matrix of coefficients associated to the linear system specified in Lemma 3, whose columns are the unknowns $g_1, \ldots, g_{n/2}, \ldots, g_{N-n/2}, \ldots, g_N$, is the following:

$$\begin{pmatrix}
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\lambda^p_{i,n/2+1} & \cdots & \lambda^p_{i,n} & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\
0 & \cdots & 0 & \lambda^p_{i,n/2+1} & \cdots & \lambda^p_{i,n} & \cdots & 0 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & \lambda^p_{i,n/2+1} & \cdots & \lambda^p_{i,n} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\lambda^q_{j,1} & \cdots & 0 & \lambda^q_{j,2} & \cdots & 0 & \cdots & \lambda^q_{j,n/2} & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & \lambda^q_{j,1} & 0 & \cdots & \lambda^q_{j,2} & \cdots & 0 & \cdots & \lambda^q_{j,n/2} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots
\end{pmatrix}.$$

We restrict our attention to the following $mn/2 \times N$ submatrix containing the equations $\mathrm{Coeff}(p_i, y_j)$ only, for $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n/2\}$:

$$\begin{pmatrix}
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\lambda^p_{i,n/2+1} & \cdots & \lambda^p_{i,n} & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\
0 & \cdots & 0 & \lambda^p_{i,n/2+1} & \cdots & \lambda^p_{i,n} & \cdots & 0 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & \lambda^p_{i,n/2+1} & \cdots & \lambda^p_{i,n} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots
\end{pmatrix},$$

We see that due to its particular shape it has rank N if there exists an $n/2 \times n/2$ invertible submatrix in the following $m \times n/2$ matrix:

$$\begin{pmatrix}
\lambda^p_{1,n/2+1} & \lambda^p_{1,n/2+2} & \cdots & \lambda^p_{1,n} \\
\lambda^p_{2,n/2+1} & \lambda^p_{2,n/2+2} & \cdots & \lambda^p_{2,n} \\
\cdots & \cdots & \cdots & \cdots \\
\lambda^p_{m,n/2+1} & \lambda^p_{m,n/2+2} & \cdots & \lambda^p_{m,n}
\end{pmatrix}. \tag{6}$$

this is, if the matrix (6) is of maximum rank. Since the coefficients of the matrix are uniformly random, the probability that a $m \times n/2$ matrix has maximum rank is, according to [7], precisely

$$\frac{(1-\frac{1}{q})\ldots(1-\frac{1}{q^m})}{(1-\frac{1}{q})\ldots(1-\frac{1}{q^{m-n/2}})} = \frac{\gamma_q(m)}{\gamma_q(m-n/2)}.$$

□

Considering that the shape of A is of the restricted form we assumed with probability $\gamma_q(n/2)$ and that the system above can be solved successfully with probability $\frac{\gamma_q(m)}{\gamma_q(m-n/2)}$, the following theorem sums up the results of this section:

Theorem 3. *Let $q > 2$. There is a randomized polynomial-time algorithm solving* HSP_q *in:*

$$\mathcal{O}(n^{2\omega}),$$

where $2 \leq \omega \leq 3$ is the linear algebra constant, and with success probability

$$\frac{\gamma_q(n/2)\gamma_q(m)}{\gamma_q(m-n/2)}.$$

The success probability of our algorithm can be asymptotically approximated by $1 - 1/q$.

Proof. The algorithm to solve HSP_q is the following:

Input: $p_1, \ldots, p_m, q_1, \ldots, q_m \in \mathbb{F}_q[\mathbf{x}]$ of degree $d \geq 3$.
Construct the linear system of Lemma 3.
Solve it.
Return this solution.

Taking into account that $\gamma_q(n) = \prod_{i=1}^{n} \left(1 - \frac{1}{q^i}\right)$ and that $\lim_{n\to\infty} \gamma_q(n) = 1 - \frac{1}{q} + \mathcal{O}\left(\frac{1}{q^2}\right)$,

$$\lim_{n\to\infty} \frac{\gamma_q(n/2)\gamma_q(m)}{\gamma_q(m-n/2)} = 1 - \frac{1}{q} + \mathcal{O}\left(\frac{1}{q^2}\right)$$

follows, and so the asymptotic success probability of our algorithm increases as we increase q. □

3.3 Experimental Results

We report here our experimental results for HSP_q, with $q > 2$, obtained with the algorithm of Theorem 3. We have implemented the algorithm using the Magma software [6]. In the tables below, `NextPrime`(k) is the Magma function that outputs the least prime number greater than k. Also, $\mathrm{Time}_{\mathrm{gen}}$ is the timed needed to generate the instances, and Time is the time spent solving the linear system. Finally, $N = n^2/4$ is the number of unknowns in the linear system.

$d = 3$					
n	q	N	$\mathrm{Time}_{\mathrm{gen}}$	Time	Memory
10	3	25	1 s	0.00 s.	13MB
12	3	36	2 s	0.00 s.	12MB
20	3	100	77.6 s	0.01 s	323MB
10	NextPrime(2^{16})	25	1 s	0.00 s	11 MB
12	NextPrime(2^{16})	36	4 s	0.00 s	12 MB
20	NextPrime(2^{16})	100	244.7 s	0.03 s	77MB

$d = 4$					
n	q	N	Time	Time	Memory
10	3	25	4 s	0.00 s	12MB
12	3	36	30 s	0.00 s	12MB
10	NextPrime(2^{16})	25	18 s	0.0 s	22MB
12	NextPrime(2^{16})	36	107 s	0.0 s	22MB
20	NextPrime(2^{16})	100	5154.050 s	0.02 s	300MB

As expected from Theorem 3 the algorithm is very efficient. Note that even for small q, all experiments performed succeeded as the probability of obtaining sufficiently many linearly independent linear equations, $\gamma_q(m)/\lambda_q(m - n/2)$, tends to 1 very quickly even for small values of q. Note that the running time of our algorithm is clearly dominated by the time spent in generating the instance. This is done in polynomial time[5] , so we can infer from the experiments that the algorithm runs in polynomial time, which is coherent with the theoretical results obtained.

[5] Note that the generation of the instance is rather slow in practice, probably due to a non-optimal implementation of the `Evaluate` function in Magma for symbolic polynomials.

4 An Efficient Algorithm for Solving HSP_2

We consider in this part the special case of HSP_2. As in Proposition 5, we can model HSP_2 by a set of algebraic equations. However, the system for $q = 2$ will have a different structure than $\mathrm{Sys}_{\mathrm{HSP}_q}$. In particular, it is no longer possible to extract linear equations (this is due to the field equations). As a consequence we have to adopt a different strategy for solving HSP_2. First we particularize the modeling for HSP_2:

Proposition 6. *Let* $\big(\mathbf{p} = (p_1, \ldots, p_m), \mathbf{q} = (q_1, \ldots, q_m)\big) \in \mathbb{F}_2[\boldsymbol{x}]^m \times \mathbb{F}_2[\boldsymbol{x}]^m$ *be a degree-d instance of* HSP_2. *Let* $\big(y_1, \ldots, y_{n/2}\big)$ *be variables and* G *be a formal matrix of size* $n/2 \times n/2$ *and* $N = n^2/4$. *We consider the system:*

$$\mathrm{Sys}_{\mathrm{HSP}_2} = \big\{\mathrm{Coeff}(p_i, t), \mathrm{Coeff}(q_j, t) \mid \forall i, j \in \{1, \ldots, m\}, \forall t \in M\big(\mathbb{F}_2[y_1, \ldots, y_{n/2}]\big)\big\} \tag{7}$$

where $\mathrm{Coeff}(p_i, t)$ *denotes the coefficient of* $t \in \mathrm{M}(\mathbb{F}_2[g_1, \ldots, g_N])$ *in* $p_i\big((y_1, \ldots, y_{n/2}) \cdot (I|G)\big)$, *and* $\mathrm{Coeff}(q_j, t)$ *the coefficient of* $t \in \mathrm{M}(\mathbb{F}_2[g_1, \ldots, g_N])$ *in* $q_j\big((y_1, \ldots, y_{n/2}) \cdot (G^T|I)\big)$. $\mathrm{Sys}_{\mathrm{HSP}_2}$ *is a system of at most*
$2m\left[\binom{n/2}{1} + \binom{n/2}{2} + \ldots + \binom{n/2}{d}\right]$ *algebraic equations over* $\mathbb{F}_2$ *in* N *variables (the entries of* G*).*

Let $A \subset \mathbb{F}_2^n$ *be a vector subspace of dimension* $n/2$. *If* A *is a solution of of* HSP_2 *on* $(\mathbf{p}, \mathbf{q}) \in \mathbb{F}_2[\boldsymbol{x}]^m \times \mathbb{F}_2[\boldsymbol{x}]^m$, *then* A *admits with probability* $\gamma_2(n/2)$ *a basis in systematic form whose components vanish all the equations* $\mathrm{Sys}_{\mathrm{HSP}_2}$.

Proof. Direct application of Proposition 5 and the fact that

$$\#(\mathrm{M}(\mathbb{F}_2[y_1, \ldots, y_{n/2}]) = \binom{n/2}{1} + \ldots + \binom{n/2}{d}.$$

□

Note that the equations of $\mathrm{Sys}_{\mathrm{HSP}_2}$ are of degree d with high probability for big enough parameters. Indeed, let G be a formal matrix of size $n/2 \times n/2$ and recall that $\mathrm{Sys}_{\mathrm{HSP}_2}$ is obtained from the coefficients of:

$$\begin{cases} p_i\left(y_1, \ldots, y_{n/2}, \sum_{t=0}^{n/2-1} g_{tn/2+1} y_{t+1}, \ldots, \sum_{t=0}^{n/2-1} g_{tn/2+n/2} y_{t+1}\right), \forall i, 1 \le i \le m, \\ q_j\left(\sum_{t=1}^{n/2} g_t y_t, \ldots, \sum_{t=1}^{n/2} g_{n/2(n/2-1)+t} y_t, y_1, \ldots, y_{n/2}\right), \forall j, 1 \le j \le m. \end{cases}$$

Since we reduce modulo the field equations, the coefficient of a linear term is obtained from the linear terms of p_i and q_j but also from the coefficients of higher degree terms reduced modulo the field equations. So, expect with a high probability that $\mathrm{Sys}_{\mathrm{HSP}_2}$ has no linear equation.

However, although $\mathrm{Sys}_{\mathrm{HSP}_2}$ is non-linear it is greatly overdetermined. Thus we can expect that computing a Gröbner basis of $\mathrm{Sys}_{\mathrm{HSP}_2}$ can still be done efficiently.

4.1 Experimental results and Interpretation

The goal of this part is to show that $\mathrm{Sys}_{\mathrm{HSP}_2}$ is indeed much easier to solve than a semi-regular system of the same size. Recall that if a system is semi-regular, its degree of regularity is given by the first non-positive coefficient of the power series specified in (2).

We report experiments run on a 2.93 GHz Intel PC with 128 Gb. of RAM with the Magma software [6] (V2.19-1) for the most disadvantageous choice of parameters (this is, $m = n$). We recall that Magma implements the F_4 algorithm ([13]) for computing Gröbner basis.

$d=3$						
n	N	$\mathrm{U}_{\mathrm{eqs}}$	$d_{\mathrm{reg}}^{\mathrm{sg}}$	d_{reg}	Time	Memory
8	16	224	4	3	1 s	17MB
10	25	500	5	3	1 s	20MB
12	36	984	5	3	2 s	55MB
14	49	1764	5	4	136 s	3Gb
16	64	2944	6	4	2.30 min	8GB
18	81	4725	7	4	2h20	80GB

$d=4$						
n	N	$\mathrm{U}_{\mathrm{eqs}}$	$d_{\mathrm{reg}}^{\mathrm{sg}}$	d_{reg}	Time	Memory
8	16	240	6	4	1 s	20MB
10	25	600	6	4	1 s	50MB
12	36	1344	7	5	38 s	840MB
14	49	2744	8	5	66 min	8GB

The notation used in the table is the following: n the number of variables of the public polynomials, $N = n^2/4$ is the number of unknowns of the system in proposition 6, $\mathrm{U}_{\mathrm{eqs}}$ is the upper bound on the number of equations as specified in proposition 6, d_{reg} is the degree of regularity observed in practice, and $d_{\mathrm{reg}}^{\mathrm{sg}}$ is the theoretical degree of regularity treating the system as if it was semi-regular. The first thing we observe is that the number of equations of our system coincides with the upper bound for the maximum number of possible equations. The experiments show that solving these systems is easier than if they were random: the degree of regularity observed in practice is indeed lower than the expected one, which suggests that there is an underlying structure that can be exploited. Furthermore, the degree of regularity stays bounded, so we conjecture the following:

Conjecture 1. The degree of regularity is bounded above by $d+1$.

If our conjecture is true, there is a randomized polynomial-time algorithm for HSP_2 as follows:

Theorem 4. *Assuming Conjecture 1, there is a randomized polynomial-time algorithm (the computation of a Gröbner basis) solving degree-d instances of* HSP_2 *with a complexity of*

$$\mathcal{O}(n^{2\omega(d+1)}),$$

where $2 \leq \omega \leq 3$ *is the linear algebra constant, and success probability* $\gamma_2(n/2)$.

5 Structural Low Degree Equations

The goal of this part is to provide theoretical arguments supporting Conjecture 1. That is, we show that the system of algebraic equations of Proposition 6 has a very particular structure. We prove that suitable linear combinations of the equations will lead to equations of a lower degree. This is actually the first computation performed by a Gröbner basis algorithm on the system of Proposition 6. As a consequence, solving the system of degree d-equations from Proposition 6 reduces to solve a system with equations of degree d and degree $d-1$. This is typically a behaviour which is not occurring in a random (i.e. semi-regular) system of equations and so it is a first step towards proving Conjecture 1.

Let $(p,q) \in \mathbb{F}_2[\mathbf{x}] \times \mathbb{F}_2[\mathbf{x}]$ be polynomials of degree d such that p vanishes on a vector subspace $A \subset \mathbb{F}_2^n$ of dimension $n/2$ and q vanishes on the orthogonal space $A^\perp$. Let $N_d = \binom{n/2}{d}$. We order lexicographically the monomials of degree d in the rings

$$\mathbb{F}_2\left[y_1,\ldots,y_{n/2}\right], \mathbb{F}_2\left[x_{n/2+1},\ldots,x_n\right], \text{ and } \mathbb{F}_2\left[x_1,\ldots,x_{n/2}\right].$$

We then denote by $t_1 < \ldots < t_{N_d}, m_1 < \ldots < m_{N_d}$ and ${m^\perp}_1 < \ldots < {m^\perp}_{N_d}$ the respective monomials in ascending order. This way we can write

$$\begin{cases} p &= \alpha_1 m_1 + \ldots + \alpha_{N_d} m_{N_d} + \tilde{p},\ \alpha_1,\ldots,\alpha_{N_d} \in \mathbb{F}_2,\ \tilde{p} \in \mathbb{F}_2\left[\mathbf{x}\right] \setminus \mathrm{M}_d(\mathbb{F}_2[x_{\frac{n}{2}+1},\ldots,x_n]), \\ q &= \beta_1 {m^\perp}_1 + \ldots + \beta_{N_d} {m^\perp}_{N_d} + \tilde{q},\ \beta_1,\ldots,\beta_{N_d} \in \mathbb{F}_2,\ \tilde{q} \in \mathbb{F}_2\left[\mathbf{x}\right] \setminus \mathrm{M}_d(\mathbb{F}_2[x_1,\ldots,x_{n/2}]). \end{cases} \tag{8}$$

Recall that the notation $\mathrm{Coeff}(p,t)$ refers to the coefficient of $t \in \mathrm{M}(\mathbb{F}_2[g_1,\ldots,g_N])$ occurring in $p\big((y_1,\ldots,y_{n/2})(I|G)\big)$, and $\mathrm{Coeff}(q,t)$ refers to the coefficient of $t \in \mathrm{M}(\mathbb{F}_2[g_1,\ldots,g_N])$ occurring in

$p\big((y_1,\ldots,y_{n/2})(G^T|I)\big)$. We will denote by $\mathrm{Coeff}(p,t)^{(d)}$ the homogeneous component of degree d of $\mathrm{Coeff}(p,t)$.

Let $t \in \mathrm{M}(\mathbb{F}_2[g_1,\ldots,g_N])$, and we can deduce from (5) that $\mathrm{Coeff}(\tilde{p},t)^{(d)} = 0 = \mathrm{Coeff}(\tilde{q},t)^{(d)}$. Then, the homogeneous component of degree d of $\mathrm{Coeff}(p,t)$ (resp. $\mathrm{Coeff}(q,t)$) is equal to the sum of the contributions with terms of degree d that each monomial m_i (resp. $m_j^{\perp}$) present in (8) makes to, respectively, $\mathrm{Coeff}(p,t)^{(d)}$ and $\mathrm{Coeff}(q,t)^{(d)}$, this is,

$$\mathrm{Coeff}(p,t)^{(d)} = \alpha_1\mathrm{Coeff}(m_1,t)^{(d)} + \ldots + \alpha_{N_d}\mathrm{Coeff}(m_{N_d},t)^{(d)},$$
$$\mathrm{Coeff}(q,t)^{(d)} = \beta_1\mathrm{Coeff}(m^{\perp}{}_1,t)^{(d)} + \ldots + \beta_{N_d}\mathrm{Coeff}(m^{\perp}{}_{N_d},t)^{(d)}.$$

The fact that G and G^T have the same entries (in different positions) and are involved in the evaluations of $p_i((y_1,\ldots,y_{n/2})(I|G))$ and $q_j((y_1,\ldots,y_{n/2})(G^T|I))$ produces certain relations between expressions of the form $\mathrm{Coeff}(m_i,t_1)^{(d)}$ and expressions of the form $\mathrm{Coeff}(m^{\perp}{}_j,t_2)^{(d)}$ for appropriate $t_1,t_2 \in \mathrm{M}_d(\mathbb{F}_2[y_1,\ldots,y_{n/2}])$. These relations are detailed in the following result:

Proposition 7. *Let $(p,q) \in \mathbb{F}_2[\boldsymbol{x}] \times \mathbb{F}_2[\boldsymbol{x}]$ be polynomials of degree d such that p vanishes on a vector subspace $A \subset \mathbb{F}_2^n$ of dimension $n/2$ and q vanishes on the orthogonal space $A^{\perp}$. Let $N_d = \binom{n/2}{d}$. For all $i,j \in \{1,\ldots,N_d\}$, it holds that:*

$$\mathrm{Coeff}\big(m_i,t_j\big)^{(d)} = \mathrm{Coeff}\big(m^{\perp}{}_j,t_i\big)^{(d)},$$

where $t_1 < \ldots < t_{N_d}, m_1 < \ldots < m_{N_d}$, and $m^{\perp}{}_1 < \ldots < m^{\perp}{}_{N_d}$ are ordered increasingly in the sets of monomials $\mathrm{M}_d(\mathbb{F}_2\left[y_1,\ldots,y_{n/2}\right]), \mathrm{M}_d(\mathbb{F}_2\left[x_{n/2+1},\ldots,x_n\right])$, and $\mathrm{M}_d(\mathbb{F}_2\left[x_1,\ldots,x_{n/2}\right])$ respectively. Also, $\mathrm{Coeff}(m_i,t_j)$ (resp. $\mathrm{Coeff}(m^{\perp}{}_j,t_i)$) denotes the coefficient of $t_i \in \mathrm{M}_d(\mathbb{F}_2\left[y_1,\ldots,y_{n/2}\right])$ (resp. $t_j \in \mathrm{M}_d(\mathbb{F}_2\left[y_1,\ldots,y_{n/2}\right])$) in $m_i\big((y_1,\ldots,y_{n/2})(I|G)\big)$ (resp. $m^{\perp}{}_j\big((y_1,\ldots,y_{n/2})(G^T|I)\big)$). Finally, the expression $\mathrm{Coeff}\big(m_i,t_j\big)^{(d)}$ (resp. $\mathrm{Coeff}\big(m^{\perp}{}_j,t_i\big)^{(d)}$) denotes the homogeneous component of degree d of $\mathrm{Coeff}\big(m_i,t_j\big)$ (resp. $\mathrm{Coeff}\big(m^{\perp}{}_j,t_i\big)$).

Proof. Given $i,j \in \{1,\ldots,N_d\}$, we have that $t_i = y_{i_1}y_{i_2}\ldots y_{i_d}, t_j = y_{j_1}y_{j_2}\ldots y_{j_d}$, for some $i_1,i_2,\ldots,i_d$, $j_1,j_2,\ldots,j_d \in \{1,\ldots,n/2\}$. On the one hand:

$$\mathrm{Coeff}(m_i,t_j)^{(d)} = \mathrm{Coeff}\big(\prod_{k=1}^{d} x_{i_k+n/2}, y_{j_1}y_{j_2}\ldots y_{j_d}\big)^{(d)}.$$

Observing the system (5), this is the coefficient of $y_{j_1}y_{j_2}\ldots y_{j_d}$ in the product

$$\prod_{k=1}^{d}\sum_{\ell=0}^{n/2-1} g_{i_k+\ell n/2}y_{\ell+1},$$

which, after expanding it, equals

$$\sum_{\pi \text{ permutation over } \{1,\ldots,d\}}\ \prod_{k=1}^{d} g_{i_k+(j_{\pi(k)}-1)n/2}. \tag{9}$$

On the other hand,

$$\mathrm{Coeff}(m^{\perp}{}_j,t_i)^{(d)} = \mathrm{Coeff}(\prod_{k=1}^{d} x_{j_k}, y_{j_1}y_{j_2}\ldots y_{j_d})^{(d)}.$$

Again, observing the system (5), this is the coefficient of $y_{i_1}y_{i_2}\dots y_{i_d}$ in the product

$$\prod_{k=1}^{d}\sum_{\ell=1}^{n/2} g_{(j_k-1)n/2+\ell}y_\ell$$

which, again after expanding it, equals

$$\sum_{\pi \text{ permutation over } \{1,\dots,d\}} \prod_{k=1}^{d} g_{(j_k-1)n/2+i_{\pi(k)}}. \tag{10}$$

Now (10) and (9) clearly coincide since

$$\sum_{\pi \text{ permutation over } \{1,\dots,d\}} \prod_{k=1}^{d} g_{(j_k-1)n/2+i_{\pi(k)}} = \sum_{\pi \text{ permutation over } \{1,\dots,d\}} \prod_{k=1}^{d} g_{i_{\pi(k)}+\left(j_{\pi^{-1}(\pi(k))}-1\right)n/2} =$$

$$\sum_{\pi^{-1} \text{ permutation over } \{1,\dots,d\}} \prod_{k=1}^{d} g_{i_k+\left(j_{\pi^{-1}(k)}-1\right)n/2}.$$

□

We can use this proposition to identify the suitable linear combinations that are of degree $d-1$:

Theorem 5. *Let the notations be as in Proposition 7. There exist $i,j \in \{1,\dots,N_d\}$ such that the equation*

$$\mathrm{Coeff}(p,t_j)+\mathrm{Coeff}(q,t_i)+\sum_{\{k\neq i|\alpha_k\neq 0\}}\mathrm{Coeff}(q,t_k)+\sum_{\{\ell\neq j|\beta_\ell\neq 0\}}\mathrm{Coeff}(p,t_\ell)$$

is of degree $d-1$.

Proof. Denote by i,j the smallest indexes such that $\alpha_i,\beta_j\neq 0$. Using Proposition 7,

$$\mathrm{Coeff}(m_i,t_j)^{(d)}=\mathrm{Coeff}(m^{\perp}{}_j,t_i)^{(d)}.$$

Now, for every $k\neq i$ such that $\alpha_k\neq 0$ and for all $\ell\neq j$ such that $\beta_\ell\neq 0$, using Proposition 7 we get the following equalities:

$$\mathrm{Coeff}(m_k,t_j)^{(d)}=\mathrm{Coeff}(m^{\perp}{}_j,t_k)^{(d)},\ \mathrm{Coeff}(m^{\perp}{}_\ell,t_i)^{(d)}=\mathrm{Coeff}(m_i,t_\ell)^{(d)},\ \text{and}$$

$$\mathrm{Coeff}(m^{\perp}{}_\ell,t_k)^{(d)}=\mathrm{Coeff}(m_k,t_\ell)^{(d)}.$$

Now adding up both the left-hand side and the right-hand side of all the four equalities, we obtain

$$\begin{aligned}&\mathrm{Coeff}(m_i,t_j)^{(d)}+\sum_{\{k\neq i|\alpha_k\neq 0\}}\mathrm{Coeff}(m_k,t_j)^{(d)}+\sum_{\{\ell\neq j|\beta_\ell\neq 0\}}\mathrm{Coeff}(m^{\perp}{}_\ell,t_i)^{(d)}+\\&+\sum_{\{\ell\neq j,k\neq i|\alpha_k\neq 0,\beta_\ell\neq 0\}}\mathrm{Coeff}(m^{\perp}{}_\ell,t_k)^{(d)}+\mathrm{Coeff}(m^{\perp}{}_j,t_i)^{(d)}+\sum_{\{k\neq i|\alpha_k\neq 0\}}\mathrm{Coeff}(m^{\perp}{}_j,t_k)^{(d)}+\\&+\sum_{\{\ell\neq j|\beta_l\neq 0\}}\mathrm{Coeff}(m_i,t_\ell)^{(d)}+\sum_{\{\ell\neq j,k\neq i|\alpha_k\neq 0,\beta_\ell\neq 0\}}\mathrm{Coeff}(m_k,t_\ell)^{(d)}=0.\end{aligned}$$

The left-hand side of this equality is the homogeneous component of degree d of

$$\mathrm{Coeff}(p, t_j) + \mathrm{Coeff}(q, t_i) + \sum_{\{k \neq i | \alpha_k \neq 0\}} \mathrm{Coeff}(q, t_k) + \sum_{\{\ell \neq j | \beta_\ell \neq 0\}} \mathrm{Coeff}(p, t_\ell)$$

which means that we cancelled out the terms of degree d and so the required equation is of degree $d-1$. □

This result can be used to generate low-degree equations:

Corollary 1 *Let* $\big(\mathbf{p} = (p_1, \ldots, p_m), \mathbf{q} = (q_1, \ldots, q_m)\big) \in \mathbb{F}_2[\boldsymbol{x}]^m \times \mathbb{F}_2[\boldsymbol{x}]^m$ *be a degree-d instance of* HSP_2. *We can easily generate* $\mathcal{O}(m^2)$ *equations of degree* $d-1$. *These equations are linear combinations of the degree-d equations of* $\mathrm{Sys}_{\mathrm{HSP}_2}$.

Proof. We apply simply Theorem 5 to each pair of polynomials $(p_i, q_j) \in \mathbb{F}_2[\mathbf{x}] \times \mathbb{F}_2[\mathbf{x}]$. From the proof of Theorem 5, it is clear that these equations are linear combinations of the equations from $\mathrm{Sys}_{\mathrm{HSP}_2}$. □

To conclude this part, we include below experimental results about the number of equations of degree $d-1$ generated as Corollary 1 which are linearly independent. In the table, we denote by $\#\mathrm{eqs}_{\mathrm{pr}}$ the number of linearly independent equations obtained in practice and by $\#\mathrm{eqs}_{\mathrm{th}}$ the maximum number of linearly independent equations that can be obtained, which is m^2.

	$d=3$		$d=4$	
	$\#\mathrm{eqs}_{\mathrm{pr}}$	$\#\ \mathrm{eqs}_{\mathrm{th}}$	$\#\mathrm{eqs}_{\mathrm{pr}}$	$\#\ \mathrm{eqs}_{\mathrm{th}}$
$m=n=10$	99	100	71	100
$m=n=12$	144	144	144	144
$m=n=14$	196	196	196	196
$m=n=16$	256	256	256	256

We observe that the behaviour is unstable for small values of the parameters. This is partially due to the fact that if a polynomial p_i (resp. q_j) does not have terms of degree d in $\mathbb{F}_2[x_{n/2+1}, \ldots, x_n]$ (resp. $\mathbb{F}_2[x_1, \ldots, x_{n/2}]$), then we do not get equations of degree $d-1$ applying Theorem 5 to the pair (p_i, q_k) for all $k \in \{1, \ldots, m\}$ (resp. from the pair (p_k, q_j) for all $k \in \{1, \ldots, m\}$). This happens with probability $1/2^{\binom{n/2}{d}}$, which is not too small for low parameters. So, for small parameters it is possible that we obtain a number of equations of degree $d-1$ lower than m^2. However, if this is the case there still are equations of degree $d-1$ (or lower) produced by the terms of degree $d-1$ (or lower). We see that the behavior becomes stable for big enough values of the parameters m, n obtaining as many equations of degree $d-1$ as possible, this is, m^2.

6 Conclusions

In this paper we presented a very efficient attack for HSP_q with $q > 2$. Since the asymptotic probability of success of this algorithm is $1 - 1/q$, the quantum money scheme extended to $\mathbb{F}_q$ is badly broken for big q. We also provided some experimental and theoretical arguments that support the conjecture that HSP_2 can be solved in polynomial time.

References

1. Scott Aaronson and Paul Christiano. Quantum money from hidden subspaces. In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 41–60, 2012.
2. Magali Bardet, Jean-Charles Faugère, and Bruno Salvy. On the complexity of Gröbner basis computation of semi-regular overdetermined algebraic equations. In *Proc. of International Conference on Polynomial System Solving (ICPSS)*, pages 71–75, 2004.
3. Magali Bardet, Jean-Charles Faugère, and Bruno Salvy. On the Complexity of the F5 Grb̈ner basis Algorithm. *Journal of Symbolic Computation*, pages 1–24, September 2014. 24 pages.
4. Magali Bardet, Jean-Charles Faugère, Bruno Salvy, and Bo-Yin Yang. Asymptotic behaviour of the degree of regularity of semi-regular polynomial systems. In *Proc. of MEGA 2005, Eighth International Symposium on Effective Methods in Algebraic Geometry*, 2005.
5. Charles H. Bennett, Gilles Brassard, Seth Breidbard, and Stephen Wiesner. Quantum cryptography, or unforgeable subway tokens. In *Proceedings of CRYPTO*, pages 267–275, 1982.
6. Wieb Bosma, John J. Cannon, and Catherine Playoust. The Magma algebra system I: The user language. *Journal of Symbolic Computation*, 24(3-4):235–265, 1997.
7. Richard P. Brent and Brendan D. McKay. Determinants and rank of random matrices over $\mathbb{Z}_m$. *Discrete Math.*, 66:35–50, 1987.
8. Bruno Buchberger. *Ein Algorithmus zum Auffinden der Basiselemente des Restklassenringes nach einem nulldimensionalen Polynomideal*. PhD thesis, University of Innsbruck, 1965.
9. Bruno Buchberger. Bruno buchberger's phd thesis 1965: An algorithm for finding the basis elements of the residue class ring of a zero dimensional polynomial ideal. *J. Symb. Comput.*, 41(3-4):475–511, 2006.
10. Bruno Buchberger. Comments on the translation of my phd thesis. *J. Symb. Comput.*, 41(3-4):471–474, 2006.
11. Nicolas Courtois and Willi Meier. Algebraic attacks on stream ciphers with linear feedback. In Eli Biham, editor, *Advances in Cryptology - EUROCRYPT 2003, International Conference on the Theory and Applications of Cryptographic Techniques, Warsaw, Poland, May 4-8, 2003, Proceedings*, volume 2656 of *Lecture Notes in Computer Science*, pages 345–359. Springer, 2003.
12. Edward Farhi, David Gosset, Avinatan Hassidim, Andrew Lutomirski, and Peter W. Shor. Quantum money from knots. pages 276–289, 2012.
13. Jean-Charles Faugère. A new efficient algorithm for computing Gröbner bases (F_4). *Journal of Pure and Applied Algebra*, 139:61–88, 1999.
14. Jean-Charles Faugère. A new efficient algorithm for computing Gröbner bases without reduction to zero (F_5). In ACM Press, editor, *International Symposium on Symbolic and Algebraic Computation-ISAAC 2002*, pages 75–83, 2002.
15. Jean-Charles Faugère and Antoine Joux. Algebraic cryptanalysis of hidden field equation (HFE) cryptosystems using gröbner bases. In Dan Boneh, editor, *Advances in Cryptology - CRYPTO 2003, 23rd Annual International Cryptology Conference, Santa Barbara, California, USA, August 17-21, 2003, Proceedings*, volume 2729 of *Lecture Notes in Computer Science*, pages 44–60. Springer, 2003.
16. Jean-Charles Faugère and Ludovic Perret. Polynomial equivalence problems: Algorithmic and theoretical aspects. In *Advances in Cryptology - EUROCRYPT 2006, 25th Annual International Conference on the Theory and Applications of Cryptographic Techniques, St. Petersburg, Russia, May 28 - June 1, 2006, Proceedings*, pages 30–47, 2006.
17. Dmitry Gavinsky. Quantum money with classical verification. pages 42–52, 2012.
18. Michele Mosca and Douglas Stebila. Quantum coins. *Error-Correcting Codes, Finite Geometry and Cryptography*, 523:35–47, 2010.
19. Jacques Patarin. Hidden fields equations (HFE) and isomorphisms of polynomials (IP): two new families of asymmetric algorithms. In *Advances in Cryptology - EUROCRYPT '96, International Conference on the Theory and Application of Cryptographic Techniques, Saragossa, Spain, May 12-16, 1996, Proceeding*, pages 33–48, 1996.
20. Stephen Wiesner. Conjugate coding. *ACM SIGACT News*, 15(1):78–88, 1983.