Ce que j'ai compris jusqu'à maintenant
Début de rapport ?

Paul Mekhail
Encadrante: Sorina Ionica

April 4, 2025

1 Introduction

The MPC in the head paradigm is a new framework introduced in [7] which is in vogue recently as we saw in NIST's second round competition for post-quantum signature schemes where 5 of the 14 candidates are based on this paradigm. The general idea is to use a NP-relation $\mathcal{R}(x, w)$ to obtain a ZK-protocol in which a prover $\mathcal{P}$ convinces a verifier $\mathcal{V}$ that she knows a valid witness w for a given (public) value of x without revealing any information about x. In [6], the authors introduce a problem called subfield bilinear collision problem (SBC) which originates from the discrete logarithm problem and constructed a post-quantum signature scheme based on this problem using the MPCitH framework. The goal of this internship is to perform a cryptanalysis of this signature scheme and try to improve it.

2 Signature Scheme

2.1 SBC Problem

The problem is the following: let q be a prime power and two positive integers k, n.
Given two vectors $\vec{u}, \vec{v} \in (\mathbb{F}_{q^k})^n$, which are linearly independent overs $\mathbb{F}_q$, find two vectors $\vec{x}, \vec{y} \in (\mathbb{F}_q)^n$ such that

$$(\vec{u} \cdot \vec{x})(\vec{v} \cdot \vec{y}) = (\vec{u} \cdot \vec{y})((\vec{v} \cdot \vec{x})$$

2.2 NSBC

The authors also introduce a particular version of this problem called normalized SBC (NSBC) where $\vec{x}, \vec{y} \in (\mathbb{F}_q)^n$ and $\vec{x} = (\vec{x'}, 1, 0), \vec{y} = (\vec{y'}, 0, 1)$.

2.3 Keys and parameters estimations

The public key of this signature scheme is $(\vec{u}, \vec{v})$ and the private key is $(\vec{x}, \vec{y})$. The signature and verification protocol use the MPCitH paradigm, they are described in detail in [6] with proof of soundness, correctness and zero-knowledge. The authors estimate that a good parameter for NSBC problem to be computationally secure are $q = 2, n = 130, k = 257$ using complexity analysis of what they think is the best known cryptanalysis of such problem in [5].

3 Cryptanalysis of the scheme

One of the known attacks given by the authors is through algebraic cryptanalysis. One can model the system by taking g as the following polynomial

$$g(x_1, ..., x_n, y_1, ..., y_n) := (\vec{u} \cdot \vec{x})(\vec{v} \cdot \vec{y}) - (\vec{u} \cdot \vec{y})(\vec{v} \cdot \vec{x})$$

or if we consider the NSBC problem

$$g(x_1, .., x_{n-2}, y_1, .., y_{n-2}) := (\sum_{i=1}^{n-2} u_i x_i + u_{n-1})(\sum_{i=1}^{n-2} v_i y_i + v_n) - (\sum_{i=1}^{n-2} u_i y_i + u_n)(\sum_{i=1}^{n-2} v_i x_i + v_{n-1})$$

Then, one can perform a Weil descent on the polynomial g because $\mathbb{F}_{q^k}$ is a $\mathbb{F}_q$-vector space and obtain the polynomial system

$$\begin{aligned} g_1(x_1, \ldots, x_{n-2}, y_1, \ldots, y_{n-2}) &= 0 \\ &\vdots \\ g_k(x_1, \ldots, x_{n-2}, y_1, \ldots, y_{n-2}) &= 0 \end{aligned}$$

The most straightforward cryptanalysis of this type of systems is to perform a Gröbner basis algorithm like Faugère's F5 [4].

The authors cite [5] to argue their parameters choice, we will dive deeper into this article's results in a later section.

3.1 Gröbner basis algorithms

In the following of the paper we will denote $R = \mathbb{K}[x_1, \ldots, x_n]$ a polynomial ring on a field $\mathbb{K}$.

Gröbner basis algorithms are used to find solutions for polynomial systems using algebraic geometry tools. For a gentle introduction to this subject [3].

In [9], Lazard introduced the connection between linear algebra and Gröbner basis theory via the usage of Macaulay matrices which is defined as follows.

Definition 3.1. *Let $\prec$ be an admissible monomial ordering on a given ring R. Given a sequence of homogeneous (resp. affine) polynomials $\mathcal{F} = (f_1, \ldots, f_m) \in R$, we associate to it the Macaulay matrix of degree D (resp. $\leq D$), denoted by $Mac_{D,m}(\mathcal{F})$ (resp. $Mac_{\leq D,m}(\mathcal{F})$), and defined as follows: the columns of the matrix are indexed by the monomials of degree D (resp. $\leq D$) sorted in decreasing order from the left to the right using the defined monomial ordering. Each row of the matrix is labeled by a tag, also called signature, (u, f_i) where u is a monomial in R and $f_i \in \mathcal{F}$ such that $deg(uf_i) = D$ (resp. $deg(uf_i) \leq D$), and the polynomial uf_i is written as vector of coefficients of monomials on the row. We denote by $\widetilde{Mac}_{D,m}(\mathcal{F})$ (resp. $\widetilde{Mac}_{\leq D,m}(\mathcal{F})$) the row echelon form of $Mac_{D,m}(\mathcal{F})$ (resp. $Mac_{\leq D,m}(\mathcal{F})$) without swapping the columns to mantain the monomial order.*

The major "issue" of Gröbner basis algorithms is the reductions to 0 in the Macaulay matrix associated to a polynomial sequence $Mac_{D,m}(\mathcal{F})$, because the size of the matrices are huge, multiple millions for practical examples, reductions to 0 can cost a lot during the linear algebra process. F5 [4] was a major improvement because it provided a criterion that avoided reduction to 0 during the linear algebra part for all regular sequences.

We will introduce the F5 criterions and the notions of regularity and semi-regularity in the following.

Proposition 3.1 (General criterion [4]). *Let d_i be the total degree of f_i. $\forall j < m$, if a row of signature (u, f_j) in the matrix $\widetilde{Mac}_{D-d_j,m-1}(\mathcal{F})$ has as leading term t', then the row (t', f_m) in $Mac_{D,m}(\mathcal{F})$ (resp. $Mac_{\leq D,m}(\mathcal{F})$) will reduce to 0 i.e. is a linear combination of the predecessors.*

Proposition 3.2 (Frobenius criterion [4]). *If a row of signature (t, f_m) in $\widetilde{Mac}_{D-d_m,m}(\mathcal{F})$ has as leading term t', then the row (t', f_m) in $Mac_{D,m}(\mathcal{F})$ (resp. $Mac_{\leq D,m}(\mathcal{F})$) will reduce to 0 i.e. is a linear combination of the predecessors.*

Definition 3.2. *The Hilbert function of a homogeneous ideal $I = \langle f_1, \dots f_m \rangle$ at degree s is defined as*

$$HF_{s,m,d}(n) = dim(R/I)_s = dim((R)_s/I_s) = dim(R)_s - dim(I)_s$$

where $d = (d_1, \dots, d_m)$ and d_i is the degree of the homogeneous polynomial f_i.

The index of regularity is when this function of s is equal to Hilbert's polynomial of s. The degree of this polynomial is exactly the dimension of the ideal. We denote by $H(I)$ the index of regularity of I.

Definition 3.3 (Semi-regularity [2]). *Let $R = \mathbb{K}[x_1, \dots, x_n]$ a polynomial ring with $\mathbb{K}$ a field. A homogeneous sequence $f_1, \dots, f_m \in R$ is semi-regular if the following conditions are verified*

- $I = \langle f_1, \dots, f_m \rangle \neq R$
- *For $i \in [1; m]$, if $g_i f_i = 0$ in $R/\langle f_1, \dots, f_m \rangle$ and $deg(g_i f_i) < H(I)$ then $g_i = 0$ in $R/\langle f_1, \dots, f_m \rangle$.*

Bardet showed in [2] that F5 criterion removed reductions to 0 for semi-regular sequences until the degree D which is the degree of regularity of I which is the ideal generated by the polynomials of the system.

Theorem 3.1 (Theorem 3.2.10 in [2]). *Let $f_1, ..., f_m$ a homogeneous sequence, such that $\langle f_1, ..., f_m \rangle$ is 0-dimensional and $\prec$ an admissible graded monomial ordering. We have,*

- *If the sequence is semi-regular, then no reductions to 0 are performed during the F5-matrix algorithm until the degree of regularity $D - 1$.*
- *If there are no reductions to 0 during the F5-matrix algorithm until the degree $D-1$, and if the matrix of degree D is full rank and is the first matrix to have more rows than columns, then the sequence is semi-regular and its index of regularity is $H(I) = D$.*

In [1], the authors show that the complexity of computing the Gröbner basis of a semi-regular system $f_1, ..., f_m \in \mathbb{K}[x_1, ..., x_n]$ is $O\left(mD\binom{n+D-1}{D}^{\omega}\right)$.

However, after some tests and after applying F5-matrix algorithm on polynomial systems generated by NSBC instances, we get reductions to 0, this means that the polynomial systems are not semi-regular when we only multiply by the x variables or by the y variables. If we multiply by the x variables and at least one y variable, we get no reductions to 0 which means that the system could be semi-regular.

3.2 Bilinear systems case

Our NSBC instances are quadratic systems that are bilinear in the variables of degree 2, overdetermined and can be semi-regular.

Definition 3.4. *Let E, F, G be vector spaces on the same base field $\mathbb{K}$. Let $f : E \times F \leftarrow G$ an application. We say that f is bilinear if*

$$\forall (x, x') \in E^2, \forall (y, y') \in F^2, \forall \lambda \in \mathbb{K}$$

$$f(x + x', y) = f(x, y) + f(x', y)$$
$$f(x, y + y') = f(x, y) + f(x, y')$$
$$f(\lambda x, y) = \lambda f(x, y)$$

We will note n_x (resp. n_y) the number of variables in x (resp. number of variables in y). In our case, $n_x = n_y = (m-1)/2$ and $m = k$. As we mentioned earlier, our polynomials systems can be semi-regular, so we get no reductions to 0 during the F5 algorithm using the Frobenius criterion for boolean polynomial systems until the Macaulay matrix of degree D_{reg}.

The article [5] and the thesis [10] give a partial answer to this problem by providing a new criterion for bilinear determined systems and a complexity analysis of their algorithm (The proofs in the end of [5] are not exact, the authors recommended to refer to [10] for better proofs.).

3.2.1 Jacobian matrices and syzygies

First we will need to introduce some important notations.

- Let $\mathbb{K}$ be a field. In our case in cryptography $\mathbb{K}$ is always finite.
- $R = \mathbb{K}[x_1, ..., x_{n_x}, y_1, ..., y_{n_y}]$
- Let $f_1, ..., f_m \in R$ be a bilinear polynomial. We denote by F_i the polynomial sequence $(f_1, ..., f_i)$ and I_i the ideal spanned by this sequence $\langle F_i \rangle$.
- $jac_x(F_{i-1})$ (resp. $jac_y(F_{i-1})$) is the jacobian matrix with respect to the two subsets of variables $x_1, ..., x_{n_x}$ (resp. $y_1, ..., y_{n_y}$).
- Let M be a $l \times c$ matrix with $l > c$. We call maximal minors of M the determinants of the $c \times c$ sub-matrices of M.
- $I_{i-1} : f_i$ is the ideal spanned by $\{g \in R \mid gf_i \in I_{i-1}\}$.

We will give important results of this article before discussing them further.

Theorem 3.2 (Theorem 2 in [5])**.** *Let $i > n_x + 1$ (resp. $i > n_y + 1$) and let s be a linear combination of maximal minors of $jac_x(F_{i-1})$ (resp. $jac_y(F_{i-1})$). Then $s \in I_{i-1} : f_i$*

The Algorithms 2 will give us the signatures of reductions to 0 during F5-matrix algorithm.

Algorithm 1 Reduce

Require: A monomial ordering $\prec$ and S a set of homogeneous polynomials and q a degree.
Ensure: T a reduced set of homogeneous polynomials of degree q.
1: $M \leftarrow Macaulay_{\prec}(S, q)$.
2: $M \leftarrow RowEchelonForm(M)$.
3: **Return** T the set of polynomials corresponding to the rows of M.

The following theorem, ensures that there are no reductions to 0 with extended the F5 criterion.

Algorithm 2 BL_Criterion

Require: m bilinear polynomials $f1, ...f_m$ such that $m \leq n_x + n_y$.
$\prec$ a monomial ordering overs R.
Ensure: V a set of pairs (h, f_i) such that $h \in I_{i-1} : f_i$ and $h \notin I_{i-1}$.

$V \leftarrow \emptyset$
for i from 2 to m **do**
  if $i > n_y$ **then**
    $T \leftarrow$ **Reduce**$(MaxMinors(jac_y(F_{i-1})), n_y + 1)$
    for h in T **do**
      $V \leftarrow V \cup \{(h, f_i)\}$
    end for
  end if
  if $i > n_x$ **then**
    $T' \leftarrow$ **Reduce**$(MaxMinors(jac_x(F_{i-1})), n_x + 1)$
    for h in T' **do**
      $V \leftarrow V \cup \{(h, f_i)\}$
    end for
  end if
end for
Return V

Algorithm 3 BilinF5Critetion

Require: (t, f_i) the signature of a row.
A matrix M in row echelon form.
Ensure: $True$ if the row will reduce to 0 otherwise $False$.

if t is the leading monomial of a row of M **or** $\exists (h, f_i) \in V$ such that $LM(h) = t$ **then**
  Return $True$.
else Return $False$.
end if

Theorem 3.3 (Theorem 4 in [5]). *Let $m, n_x, n_y \in \mathbb{N}$ such that $m < n_x + n_y$. If Conjecture 1 of [5] is true, then the set of bi-regular sequences $(f_1, ..., f_m)$ contains a nonempty Zariski open set. Moreover, if $(f_1, ..., f_m)$ is a bi-regular sequence, then there are no reductions to zero with the extended F5 criterion.*

For a formal definition of a *bi-regular* sequence, we refer to definition 8 of [5].
Finally, the authors give the following corollary

Corollary 3.1 (Corollary 3 in [5]). *The arithmetic complexity of computing a Gröbner basis of a generic bilinear system $f_1, ..., f_{n_x+n_y}$ with the F5 algorithm is bounded by*

$$\mathcal{O}\left(\binom{n_x + n_y + min(n_x + 1, n_y + 1)}{min(n_x + 1, n_y + 1)}^{\omega}\right)$$

where $2 \leq \omega \leq 3$ is the linear algebra constant.

First, we can clearly see that the complexity of algorithm 2 is exponential and is what we will try to show.

Proposition 3.3. *The complexity of Algorithm 2 is*

$$\mathcal{O}\left(\left(\sum_{i=n_x}^{m}\binom{i}{n_x}n_x^{\omega} + \binom{2n_x+1}{n_x}^{\omega}\right) + \left(\sum_{i=n_y}^{m}\binom{i}{n_y}n_y^{\omega} + \binom{2n_y+1}{n_y}^{\omega}\right)\right)$$

Proof. First, if $i > n_y$ (resp. $i > n_x$), we construct that jacobian matrix $jac_y(F_{i-1})$ (resp. $jac_x(F_{i-1})$) which is of size $n_y \times n_y$ (resp. $n_x \times n_x$) for the first one and $(m-1) \times n_y$ (resp. $(m-1) \times n_x$) for the last one. We suppose here that the construction of such matrices is free. Computing the $MaxMinors(jac_y(F_{i-1}))$ (resp. $MaxMinors(jac_x(F_{i-1}))$)means that we have to compute the determinants of all the sub-matrices of size $n_y \times n_y$ (resp. $n_x \times n_x$) and for a matrix M of size $l \times c$ with $l > c$ there are $\binom{l}{c}$ sub-matrices of size $c \times c$, hence the term $\binom{i}{n_y}n_y^{\omega}$ in the sum if we suppose the complexity of computing the determinant of a square multivariate polynomial matrix of size $c \times c$ to be $\mathcal{O}(c^{\omega})$, which is not the case but let's suppose the best case ever. faudrait trouver la vraie complexité de ça pour être plus juste. For now, we have a sequence of polynomials and we apply the algorithm **Reduce** on them with degree $n_y + 1$ (resp. $n_x + 1$). In **Reduce**, we construct the Macaulay matrix of the sequence of degree $n_y + 1$ (resp. $n_x + 1$) and compute it's row echelon form. The size of Macauly matrix is bounded by $\binom{n+D}{D}$ where D is the degree and n the number variables. So the complexity of **Reduce** is $\mathcal{O}\left(\binom{n_y+n_y+1}{n_y+1}^{\omega}\right)$ (resp. $\mathcal{O}\left(\binom{n_x+n_x+1}{n_x+1}^{\omega}\right)$). Hence the second term in the sum, because we call **Reduce** at each iteration of the for loop. We finally suppose that the iteration on the elements of T and of T' is negligible in the big O notation. □

It's an asymptotic complexity which doesn't take into account the constant factor in implementations that could slow down a lot the algorithm. We also recall that in our case with NSBC, $n_x = n_y$. So we get

$$\mathcal{O}\left(2n_x^{\omega}\left(\sum_{i=n_x}^{m}\binom{i}{n_x} + \binom{2n_x+1}{n_x}^{\omega}\right)\right)$$

Since we have quadratic systems, reductions to 0 won't appear before degree 4 of the Macaulay matrix. Also, in practice we are using $q = 2$ and there's much better algorithms in the literature

for boolean systems than basic Gröbner basis algorithms like F5. The best algorithm used at the moment is CrossBred [8]. Most of the time in practical implementations, we stop at degree 4 of the Macaulay in pre-processing step. So, to know what advantage we could have if we use the extended F5 criterion, we will suppose the worst, that all of our entries will reduce to 0 at degree 4 of the Macaulay matrix, which is unlikely.

In following, we will denote $n_x + n_y$ by simply n.

If this case arises, the only computations that would be done are for extended F5 criterion. The complexity for F5-Matrix algorithm with only the Frobenius criterion 3.2 is

$$\mathcal{O}\left(\binom{n+4}{4}^{\omega}\right)$$

If we only take into consideration the big O notation and not the constants, which is not realistic at all for efficient implementations to really break a cryptographic scheme, we get

$$Adv(n,m) = 2\left(\sum_{i=n/2}^{m}\binom{i}{n/2}\left(\frac{n}{2}\right)^{\omega} + \binom{n+1}{n/2}^{\omega}\right) - \binom{n+4}{4}^{\omega} \tag{1}$$

Using the computer algebra system SageMath [11] to test numerically when using the F5 extended criterion of [5] becomes negligible compared to the computations of reductions to 0, we see experimentally that $Adv(10,12) = 2498903$, i.e. the theoretical advantage becomes negligible and it would be advantageous to simply compute reductions to 0 instead of applying the F5 extended criterion. This number grows exponentially with n and m. For those who wonder, what it would be the estimated theoretical advantage computing reductions to 0 for the parameters given by [6] $Adv(256,257)$ is a 514 bit number.

No need to say that it's useless to use the criterion as shown in [5] without adapting it to get an advantage. This can be explained by the fact that the extended criterion computes reduced polynomials of degree n_x and n_y and in our case, n_x and n_y can be too big to compute. Maybe there could be a way to compute reduced polynomials of only degree 2 for our use case if we stop Macaulay matrix computations at degree 4 with our quadratic systems.

As the authors of [5] noted at the end of their article, generic bilinear systems are regular. They didn't show that for overdetermined systems but we saw that experimentally but it's yet to be proven.

Definition 3.5 (Definition 4 of [12]). *Let $\mathcal{F} = f_1, ..., f_m$ be a semi-regular sequence of polynomials in $\mathbb{F}_2[x_1, ..., x_n]$ and let $0 \leq \gamma \leq 1$ such that $k = (1-\gamma)n$. We say that this sequence is γ-strong semi-regular if*

$$\mathcal{S}(I) = \{(a_{k+1}, ..., a_n) \in \mathbb{F}_2^{n-k} |$$

$$\{f_1(x_1, ..., x_k, a_{k+1}, ..., a_n), ..., f_m(x_1, ..., x_k, a_{k+1}, ..., a_n)\} \text{ is not semi-regular}\}$$

has cardinality $\mathcal{O}(2^{-\gamma n})$

Note pour moi-même, il faut maintenant essayer de réaliser une implémentation de ce critère en sage ou avec hpXbred et BeanPolE pour voir comment ça se comporterait avec des petits systèmes et voir quel serait l'avantage ou le désavantage sachant que la supposition que toutes les lignes se réduisent à 0 est extrêment peu probable.

3.3 Boolean systems and Crossbred

References

[1] Magali Bardet, Jean-Charles Faugère, and Bruno Salvy. "On the complexity of the F5 Gröbner basis algorithm". In: *Journal of Symbolic Computation* 70 (2015), pp. 49–70. ISSN: 0747-7171. DOI: https://doi.org/10.1016/j.jsc.2014.09.025. URL: https://www.sciencedirect.com/science/article/pii/S0747717114000935.

[2] Magali Bardet (Turrel Bardet). "Etude des systèmes algébriques surdéterminés : applications aux codes correcteurs et à la cryptographie". Thèse de doctorat dirigée par Faugère, Jean-Charles Informatique Paris 6 2004. PhD thesis. 2004, 1 vol., [XIII] –157 p. URL: http://www.theses.fr/2004PA066404.

[3] David A. Cox, John Little, and Donal O'Shea. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. 3rd. Springer Publishing Company, Incorporated, 2010. ISBN: 1441922571.

[4] Jean Charles Faugère. "A new efficient algorithm for computing Gröbner bases without reduction to zero (F5)". In: *Proceedings of the 2002 International Symposium on Symbolic and Algebraic Computation*. ISSAC '02. Lille, France: Association for Computing Machinery, 2002, 75–83. ISBN: 1581134843. DOI: 10.1145/780506.780516. URL: https://doi.org/10.1145/780506.780516.

[5] Jean-Charles Faugère, Mohab Safey El Din, and Pierre-Jean Spaenlehauer. "Gröbner bases of bihomogeneous ideals generated by polynomials of bidegree (1,1): Algorithms and complexity". In: *Journal of Symbolic Computation* 46.4 (2011), pp. 406–437. ISSN: 0747-7171. DOI: https://doi.org/10.1016/j.jsc.2010.10.014. URL: https://www.sciencedirect.com/science/article/pii/S0747717110001902.

[6] Janik Huth and Antoine Joux. *MPC in the head using the subfield bilinear collision problem*. Cryptology ePrint Archive, Paper 2023/1685. 2023. URL: https://eprint.iacr.org/2023/1685.

[7] Yuval Ishai et al. "Zero-knowledge from secure multiparty computation". In: *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*. STOC '07. San Diego, California, USA: Association for Computing Machinery, 2007, 21–30. ISBN: 9781595936318. DOI: 10.1145/1250790.1250794. URL: https://doi.org/10.1145/1250790.1250794.

[8] Antoine Joux and Vanessa Vitse. *A crossbred algorithm for solving Boolean polynomial systems*. Cryptology ePrint Archive, Paper 2017/372. 2017. URL: https://eprint.iacr.org/2017/372.

[9] D. Lazard. "Gröbner bases, Gaussian elimination and resolution of systems of algebraic equations". In: *Computer Algebra*. Ed. by J. A. van Hulzen. Berlin, Heidelberg: Springer Berlin Heidelberg, 1983, pp. 146–156. ISBN: 978-3-540-38756-5.

[10] Pierre-Jean Spaenlehauer. "Résolution de systèmes multi-homogènes et déterminantiels algorithmes - complexité - applications". Thèse de doctorat dirigée par Faugère, Jean-Charles Informatique Paris 6 2012. PhD thesis. 2012, 1 vol. (207 p.) URL: http://www.theses.fr/2012PA066467.

[11] The Sage Developers. *SageMath, the Sage Mathematics Software System (Version x.y.z)*. https://www.sagemath.org. YYYY.

[12] Damien Vidal, Sorina Ionica, and Claire Delaplace. *An analysis of the Crossbred Algorithm for the MQ Problem*. Cryptology ePrint Archive, Paper 2024/992. 2024. DOI: https://doi.org/10.62056/ak86cy7qiu. URL: https://eprint.iacr.org/2024/992.