Solving Underdetermined Systems of Multivariate Quadratic Equations revisited

Enrico Thomae and Christopher Wolf

Horst Görtz Institute for IT-security
Faculty of Mathematics
Ruhr-University of Bochum, 44780 Bochum, Germany
enrico.thomae@rub.de,chris@Christopher-Wolf.de

Abstract. Solving systems of m $\mathcal{M}$ultivariate $\mathcal{Q}$uadratic ($\mathcal{MQ}$) equations in n variables is one of the main challenges of algebraic cryptanalysis. Although the associated $\mathcal{MQ}$-problem is proven to be NP-complete, we know that it is solvable in *polynomial time* over fields of even characteristic if either $m \geq n(n-1)/2$ (*overdetermined*) or $n \geq m(m+1)$ (*underdetermined*). It is widely believed that $m = n$ has worst case complexity. Actually in the overdetermined case Gröbner Bases algorithms show a gradual decrease in complexity from $m = n$ to $m \geq n(n-1)/2$ as more and more equations are available. For the underdetermined case no similar behavior was known. Up to now the best way to deal with the case $m < n < m(m+1)$ was to randomly guess variables until $m = n$. This article shows how to smartly use additional variables and thus obtain a gradual change of complexity over even characteristics also for the underdetermined case. Namely, we show how a linear change of variables can be used to reduce the overall complexity of solving a $\mathcal{MQ}$-system with m equations and $n = \omega m$ variables for some $\omega \in \mathbb{Q}_{>1}$ to the complexity of solving a $\mathcal{MQ}$-system with only $(m - \lfloor \omega \rfloor + 1)$ equations and variables, respectively. Our algorithm can be seen as an extension of the previously known algorithm from Kipnis-Patarin-Goubin (extended version of Eurocrypt '99) and improves an algorithm of Courtois *et al.* which eliminates $\lfloor \log_2 \omega \rfloor$ variables. For small ω we also adapt our algorithm to fields of odd characteristic. We ably our result to break current instances of the Unbalanced Oil and Vinegar public key signature scheme that uses $n = 3m$ and hence $\omega = 3$.

Key words: Underdetermined Multivariate Equations, UOV Signature Scheme

1 Introduction

It is well known that algebraic equations can be an Achilles' heel for cryptographic systems. Whether stream ciphers [5, 13], hash functions [19] or block ciphers [16], they all can be expressed through a system of equations over a finite field $\mathbb{F}$ with a solution that yields the private key. For asymmetric schemes the importance is even more obvious. For example variants of McEliece [12] or

$\mathcal{M}$ultivariate $\mathcal{Q}$uadratic ($\mathcal{MQ}$) schemes such as Hidden Field Equations [11] were broken using algebraic techniques. So it is fair to say that solving systems of $\mathcal{MQ}$ equations is one of the main challenges of algebraic cryptanalysis. However, as the underlying $\mathcal{MQ}$-problem is proven to be NP-complete [14], we cannot hope to find an efficient algorithm for all instances. In particular, if the number of equations m equals the number of unknowns n, all known empirical algorithms are exponential on random instances of the $\mathcal{MQ}$-problem. Nevertheless we know that the problem becomes easy for fields of characteristic 2 if either $m \geq n(n-1)/2$ or $n \geq m(m+1)$. In the first case, we replace each monomial by a new variable and solve a linear system in $n(n-1)/2$ equations and variables. The second case is covered by an algorithm of Kipnis-Patarin-Goubin [15, Sec. 7] and will be further explored in this article.

Until now, research mainly covered the *overdetermined* case $m \geq n$. There are many algorithms like F_4, F_5 and XL that benefit of additional equations [8, 9, 10]. So for $m = n$ even guessing one or two variables can help to reduce the complexity dramatically [2]—and thus make a big difference in practice. In contrast none of the algorithms benefits in the same way of the *underdetermined* case $n > m$ (cf. Section 1.1). In particular, their complexity is exponentially linked to the number of variables. Hence, having more variables will dramatically increase their running time (and also space requirements). As finding *one* solution often suffices for cryptographic purpose, the best way of "using" more variables today, is to fix them to random values and thus receive a hard instance with $n = m$ and one solution on average. This is not very sophisticated and in a sense similar to throw away additional equations in the overdetermined case and only work with the remaining ones. This article shows how to us additional variables and hence closes the complexity gap between $n = m$ and $n \geq m(m+1)$. Our main result apply for fields of even characteristic. In section 6 we discuss a generalization to arbitrary characteristics.

1.1 Related Work

The best treatment of the overdetermined case $m \geq n$ is covered by XL or Gröbner bases algorithms like F_4 or its successor F_5. The overall complexity is well understood [1] and becomes gradually easier if more and more equations are available. In particular for $m \geq n(n-1)/2$ over $\mathbb{F}_{2^k}$ and $m \geq n(n+1)/2$ over $\mathbb{F}_{p^k}$ for p an odd prime, the overall problem can be solved in polynomial time by *Linearization*. For the underdetermined case not much is known. Basically, all research so far has centered around two cases: $n = m$ and $n \geq m(m+1)$. The first has *exponential*, the latter *polynomial* time complexity. In particular, an algorithm from Kipnis-Patarin-Goubin [15, Sec. 7] can efficiently solve the latter case in $\mathbb{F}_{2^k}$. Courtois *et al.* [6] extended this result to arbitrary fields $\mathbb{F}_{p^k}$ and showed that the problem becomes polynomial as soon as $n \geq 2^{\frac{m}{7}}(m+1)$. Furthermore they showed how to eliminate $\lfloor \log_2 \omega \rfloor$ variables and thus receive a system of $m - \lfloor \log_2 \omega \rfloor$ variables and equations (cf. Prop. 1 in [6]). We extend this result, using a tight analysis of the technique of Kipnis-Patarin-Goubin, to receive a system of $m - \lfloor \omega \rfloor + 1$ variables and equations.

1.2 Achievement and Organization

We close an important gap in understanding the underdetermined case especially for $\mathbb{F}_{2^k}$. In particular we show that there is a gradual change from *exponential* running time to *polynomial* running time if n gets larger than m. This improves the cryptanalysis of the Unbalanced Oil and Vinegar Signature scheme (UOV) [17] and therefore forces a change of parameter sets (cf. section 5).
The organization of the paper is as follows. Section 2 gives some notation. Section 3 shows how to describe the transformation of variables we are using, shortly repeats the algorithm of Kipnis-Patarin-Goubin and introduce our new algorithm. Section 4 is the most important, as it gives a theoretical analysis of the correctness of our algorithm and also the one of Kipnis-Patarin-Goubin. Section 5 gives a complexity analysis and shows that parameters of UOV have to be increased. In section 6 we adapt our algorithm to the general case $\mathbb{F}_{p^k}$ for small ω and motivate future research on this question.

2 Notation

A $\mathcal{MQ}$-system of equations over a finite field $\mathbb{F}_q$ with q elements is given by m equations $p^{(k)} = 0$ for polynomial functions $p^{(k)} : \mathbb{F}_q^n \to \mathbb{F}_q$ for $1 \leq k \leq m$ and $\gamma_{ij}^{(k)}, \beta_i^{(k)}, \alpha^{(k)} \in \mathbb{F}_q$ according to

$$p^{(k)}(x_1, \ldots, x_n) := \sum_{1 \leq i \leq j \leq n} \gamma_{ij}^{(k)} x_i x_j + \sum_{1 \leq i \leq n} \beta_i^{(k)} x_i + \alpha^{(k)} . \tag{1}$$

If we speak of *solving* such an $\mathcal{MQ}$-system, we always mean finding *one* solution. For cryptanalytic purposes, this is actually sufficient in most cases. We call $p^{(k)}$ as defined by (1) *inhomogeneous*. The *homogeneous* case consists only of terms in $x_i x_j$ for $1 \leq i \leq j \leq n$ and is thus defined by

$$p^{(k)}(x_1, \ldots, x_n) := \sum_{1 \leq i \leq j \leq n} \gamma_{ij}^{(k)} x_i x_j .$$

The corresponding $\mathcal{MQ}$-map $\mathcal{P} : \mathbb{F}_q^n \to \mathbb{F}_q^m$ is defined by $\mathcal{P} := \left(p^{(1)}, \ldots, p^{(m)}\right)^\mathsf{T}$. To ease notation, we restrict to homogeneous systems in the sequel. Note that our algorithm also works for inhomogeneous systems without introducing a homogenization variable.

Let $\pi^{(k)}$ be the coefficient vector of $p^{(k)}(x_1, \ldots, x_n)$ in lexicographic order, *i.e.*

$$\pi^{(k)} = (\gamma_{11}^{(k)}, \gamma_{12}^{(k)}, \ldots, \gamma_{1n}^{(k)}, \gamma_{22}^{(k)}, \gamma_{23}^{(k)}, \ldots, \gamma_{nn}^{(k)}).$$

Note that our algorithm also works with other monomial orderings. However, for the ease of explanation, we have fixed lexicographic ordering throughout this paper. The corresponding coefficient matrix Π is defined by $\Pi := \left(\pi^{(1)}, \ldots, \pi^{(m)}\right)^\mathsf{T}$.

3 Transformation of Variables

Let $\mathcal{P} : \mathbb{F}_q^n \to \mathbb{F}_q^m$ be an $\mathcal{MQ}$-map with m equations and $n = \omega m$ variables $x_1, \ldots, x_n$ for some $\omega \in \mathbb{Q}_{>1}$. To make parts of the arguments easier, we will sometimes change to the notation $n = m + v$ with $v = (\omega - 1)m$. The current way to find a solution of this system is to fix v variables at random [2, 3, 7] and solve the remaining system of m equations and m variables using a $\mathcal{MQ}$-solver such as F_5 or XL. Kipnis-Patarin-Goubin [15, Sec. 7] were the first who took benefit of the additional v variables and showed that the system is solvable in *polynomial time* for $n \geq m(m+1)$. In a nutshell they applied a linear transformation $S \in \mathrm{GL}_n(\mathbb{F}_q)$ of variables to obtain a new $\mathcal{MQ}$-system $\mathcal{F} := \mathcal{P} \circ S$ with coefficient matrix Φ. The transformation matrix S is calculated in polynomial time such that fixing v variables in $\mathcal{F}$ provides a linear system in the remaining m variables for fields of characteristic 2. We will investigate this approach in more detail in section 3.2.

To understand how S operates on the coefficients of $\mathcal{P}$ and $\mathcal{F}$, we introduce the transformation Σ such that $\Sigma\Pi^\intercal = \Phi^\intercal$. This transformation was previously used to determine short key variants of UOV [18].

3.1 How to Determine Σ

We can write every equation $p^{(i)}$ of $\mathcal{P}$ as a quadratic form $p^{(i)} = x^\intercal \mathfrak{P}^{(i)} x$ for $x = (x_1, \ldots, x_n)$ and a matrix $\mathfrak{P}^{(i)} \in \mathbb{F}_q^{n \times n}$ consisting of the coefficients of $p^{(i)}$. Note that this matrix is *not* symmetric if $\mathbb{F}$ is of characteristic 2. Applying the change of variables, *i.e.* $y = S^{-1}x$, we obtain a new $\mathcal{MQ}$-system $\mathcal{F}$ with $f^{(i)} = y^\intercal S^\intercal \mathfrak{P}^{(i)} S y$ for $y = (y_1, \ldots, y_n)$. The coefficients of this new map are determined by $S^\intercal \mathfrak{P}^{(i)} S =: \mathfrak{F}^{(i)}$. Or in other words $f^{(k)}(y_1, \ldots, y_n) := \sum_{1 \leq i \leq j \leq n} \widetilde{\gamma}_{ij}^{(k)} y_i y_j$ and $x_i = \sum_{p=1}^{n} s_{ip} y_p$. Comparison of coefficients in the following equation reveals an explicit formula for Σ.

$$\begin{aligned} \sum_{1 \leq i \leq j \leq n} \widetilde{\gamma}_{ij}^{(k)} y_i y_j &= \sum_{1 \leq i \leq j \leq n} \gamma_{ij}^{(k)} x_i x_j \\ &= \sum_{1 \leq i \leq j \leq n} \gamma_{ij}^{(k)} (\sum_{p=1}^{n} s_{ip} y_p)(\sum_{p=1}^{n} s_{jp} y_p) \end{aligned}$$

Let $s_{i\cdot} \in \mathbb{F}_q^n$ be the i-th row of S and $D^{ij} := s_{i\cdot}^\intercal s_{j\cdot}$ the dyadic product of the i-th and j-th row of S. Now we can express $x_i x_j$ by

$$x_i x_j = (\sum_{p=1}^{n} s_{ip} y_p)(\sum_{p=1}^{n} s_{jp} y_p) = \sum_{1 \leq l \leq n} D_{ll}^{ij} y_l^2 + \sum_{1 \leq l < p \leq n} (D_{lp}^{ij} + D_{pl}^{ij}) y_l y_p.$$

Let $I := ((a_i, b_i) \mid 1 \leq a_i \leq b_i \leq n)$ be the ordered index set of all quadratic monomials. We have chosen lexicographic order of the monomials, *i.e.*

$$I = ((1,1), (1,2), \ldots, (1,n), (2,2), (2,3), \ldots, (n,n)).$$

For (a_i, b_i), $(a_j, b_j) \in I$ we obtain $\Sigma := (\sigma_{ij})$ with

$$\sigma_{ij} := \begin{cases} s_{a_j a_i} \cdot s_{b_j b_i} & \text{for } a_i = b_i, \\ s_{a_j a_i} \cdot s_{b_j b_i} + s_{a_j b_i} \cdot s_{b_j a_i} & \text{for } a_i \neq b_i, \end{cases} \tag{2}$$

by collecting the appropriate entries of all the dyadic products. The matrix Σ, obtained by comparing coefficients, maps the coefficients of $p^{(k)}$ to the coefficients of $f^{(k)}$. Denoting $\tau := |I| = \frac{n(n+1)}{2}$ (for $q > 2$) and $\tau := |I| = \frac{n(n-1)}{2}$ (for $q = 2$) the number of monomials, this leads to

$$\Sigma \cdot \begin{pmatrix} \gamma^{(k)}_{a_1,b_1} \\ \vdots \\ \gamma^{(k)}_{a_\tau,b_\tau} \end{pmatrix} = \begin{pmatrix} \widetilde{\gamma}^{(k)}_{a_1,b_1} \\ \vdots \\ \widetilde{\gamma}^{(k)}_{a_r,b_r} \end{pmatrix}. \tag{3}$$

3.2 Algorithm of Kipnis-Patarin-Goubin

We briefly explain the algorithm of Kipnis-Patarin-Goubin for $n \geq m(m+1)$, cf. [15, Sec. 7] for details. In section 3.3 we will generalize this technique to $n \leq m(m+1)$ and show that we can force enough elements of Φ, *i.e.* coefficients of $\mathcal{F}$, to be zero, such that we obtain $(\lfloor \omega \rfloor - 1)$ linear equations. The first idea is to split the variables $y_1, \ldots, y_n$ into two sets $V := \{y_{m+1}, \ldots, y_n\}$ and $O := \{y_1, \ldots, y_m\}$. Here V denotes the set of variables we want to fix and O the set of variables we want to determine. Due to the strong connection to the Oil and Vinegar Signature Scheme, we call V the *vinegar variables* and O the *oil variables*. The aim of Kipnis-Patarin-Goubin was to find S such that most coefficients of $\mathcal{F}$ are zero and thus the new $\mathcal{MQ}$-system is easily solvable *e.g.* by Linearization. The overall idea to find such a linear transformation S efficiently is the following. First all equations of (3) are quadratic in s_{ij}. But if we fix certain elements of S at random, some of the equations become linear. Solving this linear equations enable us to fix some coefficients of $\mathcal{F}$ to zero. More precisely Kipnis-Patarin-Goubin aimed at solving the quadratic equations in s_{ij} of (3) we obtain by setting

$$\widetilde{\gamma}^{(k)}_{i,j} = 0 \text{ for } 1 \leq i, j, k \leq m, i \neq j\,. \tag{4}$$

To ease notation we label (4) by (i, j, k) or just (i, j) if we want to denote all equations $(i, j, 1)$ to (i, j, m). As all these equations are quadratic, Kipnis-Patarin-Goubin fixed the first column of S to random values. Note that regarding to (2) all monomials in equation (i, j) consists of one variable of the i-th and one variable of the j-th column of S. This means $\widetilde{\gamma}^{(k)}_{1,1}$ is fixed to a random value and equations $(1, 2)$ to $(1, n)$ become linear. $(1, 2)$ gives us m linear equations in the s_{i2} and after randomly fixing the superfluous variables, we can determine them such that $\widetilde{\gamma}^{(k)}_{1,2} = 0$ for $1 \leq k \leq m$. Now that the second column of S is determined, we obtain additional linear equations $(2, 3)$ to $(2, n)$. Using the $2m$ linear equations of $(1, 3)$ and $(2, 3)$, we can determine s_{i3}. If the first k columns of S are determined, we solve the km linear equations $(1, k+1)$ to $(k, k+1)$ to

determine the $(k+1)$-th column of S. The algorithm continues until columns 1 to m of S are determined. At each level, more and more of the equations become linear. For the last step we have to solve the linear equations $(1,m)$ to $(m-1,m)$ in the unknowns s_{1m} to s_{nm} and thus $n \geq m(m-1)$ must hold.
After this transformation we obtain m equations $1 \leq j \leq m$ of the form

$$\sum_{i=1}^{m} \beta_{i,j} y_i^2 + y_1 L_{1,j}(y_{m+1}, \ldots, y_{m+v}) + \ldots \\ + y_m L_{m,j}(y_{m+1}, \ldots, y_{m+v}) + Q_j(y_{m+1}, \ldots, y_{m+v}) \,. \tag{5}$$

The terms $L_{i,j}$ denote some linear functions in the V variables we want to fix and Q_j denotes some quadratic function in these variables. Now Kipnis-Patarin-Goubin determined $y_{m+1}, \ldots, y_{m+v}$ by Gaussian Elimination such that $L_{i,j} = 0$ for all $1 \leq i, j \leq m$. This is possible for $v \geq m^2$ and thus we obtain the condition $n \geq m(m+1)$. For fields of characteristic 2 the remaining system in (5) is linear in the O variables and can thus be easily solved. This is due to the Frobenius Homomorphism $x \mapsto x^2$ which effectively allows us to treat monomials of the form y_i^2 as linear variables.
In the next section we provide a tight analysis for $n \leq m(m+1)$ and show that solving a $\mathcal{MQ}$-system $\mathcal{P}$ with m equations and $n = \omega m$ variables is roughly as hard as solving a $\mathcal{MQ}$-system of $(m - \lfloor \omega \rfloor + 1)$ equations in $(m - \lfloor \omega \rfloor + 1)$ variables.

3.3 Tight Analysis for $n = \omega m$ and Improvement

To obtain linear equations we also fix the first column of S at random. This step is similar to Kipnis-Patarin-Goubin. As we are in the case $m < n < m(m+1)$ we cannot fulfill all equations (4) and have to adjust our strategy accordingly. In particular, we have to reduce the number of equations during the intermediate steps, *i.e.* due to a lack of variables in S we can only solve equations (i,j,k) for some fixed bound b_j and $1 \leq i < j$, $1 \leq k \leq b_j$.

The overall process is depicted in figure 1. Remember, lines of $\Phi^\intercal$ denote coefficients and columns denote polynomials $f^{(1)}$ to $f^{(m)}$. The dotted areas are of no interest for us, as all the corresponding monomials in $\mathcal{F}$ vanish after fixing the variables in V. Hence, setting them to a specific value is no use. The gray part are arbitrary values. To make the interesting part of $\Phi^\intercal$ clearer, we reordered the rows in the right block of figure 1, what is indicated by the ordered pairs labeling the rows.
Let $B_j := \{(i,j,k) \,|\, 1 \leq k \leq b_j, 1 \leq i < j\}$ be the j-th block of coefficients for which we want to gain $\widetilde{\gamma}_{i,j}^{(k)} = 0$ in Φ where b_j is some bound to be determined later (cf. section 4) and $\mathcal{V}(B_j) := (j-1)b_j$ the volume of such a block or in other words the number of zero coefficients. As depicted in figure 1 we are able to eliminate all $O \times O$ coefficients with $i \neq j$ in the first b_m columns of Φ. Solving the linear system in the $\widetilde{\gamma}_{i,i}^{(k)} y_i^2$ for $i \in \{1, \ldots, m\}$ allows us to replace a total of

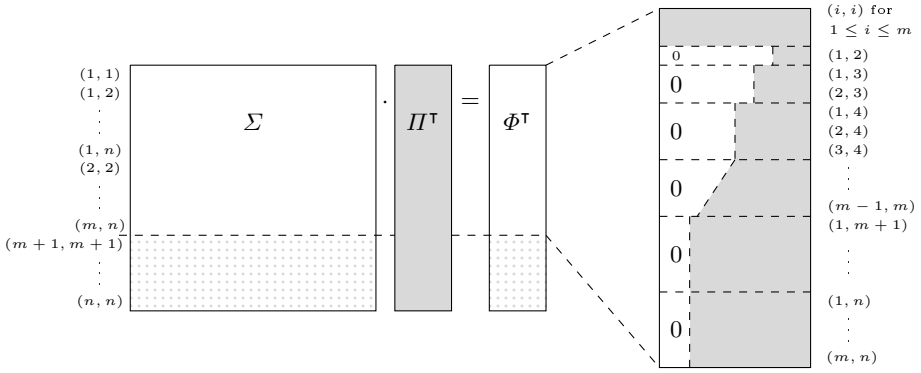

Fig. 1. Overview of coefficients in $\Phi^\intercal$ we want to fix to zero.

b_m variables in the remaining equations (see Section 4.2). This leads effectively to a new system of $(m - b_m)$ equations and variables. The crucial point is to determine the correct value of b_m. One might be inclined to choose $b_j \leq m$ maximal such that $\mathcal{V}(B_j)$ is less than the number of variables n in the j-th column of S in order to produce as much zeros in Φ as possible. However note that $\mathcal{V}(B_j)$ cannot be equal to n as we need one more variable than equations in each block, as the system is homogeneous. Hence, by having as many variables as linear equations, we only would obtain the all-zero vector. In section 4 we will show that the naive approach will not work in general, as if b_j is to large the obtained solution S will not be regular. This question did not come up in the Kipnis-Patarin-Goubin approach, as they fixed enough variables at random to trivially assure regularity of S.

Algorithm 1 High-level description of our algorithm.

1: Fix columns $m+1$ to n of S at random.
2: **for all** $i = 1 \to m$ **do**
3: **for all** $j = 1 \to i - 1$ **do**
4: Collect all linear equations $(i, j, 1)$ to (i, j, ω).
5: **end for**
6: Solve them and include their solution to S.
7: **end for**
8: Apply linear transformation S.
9: **for all** $i = 1 \to \omega$ **do**
10: Collect all linear equations $L_{1,i}$ to $L_{m,i}$.
11: **end for**
12: Solve them and derive vinegar variables.
13: Substitute ω linear equations in remaining $\mathcal{MQ}$-polynomials.

4 Equivalent Solutions S and their Impact on b_j

Up to this point our approach is a straightforward enhancement of Kipnis-Patarin-Goubin idea. This section covers the main idea of our approach and gives new insights to the theory of solving underdetermined systems of equations. Kipnis-Patarin-Goubin claimed that all the linear equations provide at least one solution in general and that S is regular with high probability. Due to the large n this is actually true for their approach. But as we want to use as many s_{ij} as possible to fix elements in Φ to zero, it is not clear at all, how many equations $\widetilde{\gamma}_{i,j}^{(k)} = 0$ we are able to solve in order to obtain a regular solution S. We use the theory of equivalent keys [20, 21] for the Unbalanced Oil and Vinegar Scheme as a toolkit to show that for every solution S there is an equivalent solution S' with a special structure. The number of variables in S' will upper bound the number of equations $\widetilde{\gamma}_{i,j}^{(k)} = 0$ that yield a regular solution.

4.1 Equivalent Solutions

First let us determine b_m and thus the number of zeros in Φ for $n \geq m(m+1)$, *i.e.* for the original algorithm of Kipnis-Patarin-Goubin.

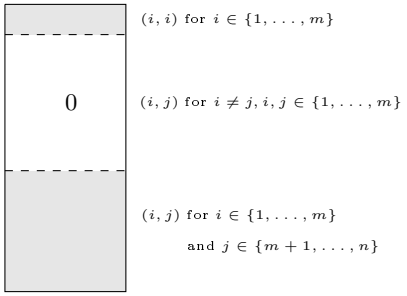

Fig. 2. Upper part of Φ^{T} for $n \geq m(m+1)$.

Denote $\mathbb{S}$ the subset of equations of $\Sigma\Pi^{\mathsf{T}} = \Phi^{\mathsf{T}}$ labeled by (i,j) for $i,j \in \{1,\ldots,m\}$ and $i \neq j$, *i.e.* the zero part in figure 2. Let S be a solution to $\mathbb{S}$. We call S' an *equivalent solution*, if it preserves the structure of Φ, *i.e.* if S' also fulfills all equations of $\mathbb{S}$. Every element of the equivalence class of such solutions solves our problem. To determine an upper bound on b_j we search for a small (linear) family of matrices $\widetilde{S}$ such that every equivalence class has a representative in this family. We call this a minimal representative. Or loosely speaking these are solutions with large fixed parts for arbitrary Π or a matrix $\widetilde{S}$ with minimal number of variables.
Obviously all equations in $\mathbb{S}$ remain zero if we map every variable $\{y_1,\ldots,y_m\}$ to itself or some permutation and any variable within $\{y_{m+1},\ldots,y_n\}$ to sums

of these variables. The only two things we are not allowed to do is mapping variables of V to variables of O as this would lead to quadratic terms in the O variables and mapping O variables to a sum of O variables, as this would also lead to quadratic terms in the O variables due to $\widetilde{\gamma}_{i,i}^{(k)} \neq 0$ for $i \in \{1, \ldots, m\}$. So if S is a solution to $\mathbb{S}$ then $S\Omega^{-1}$ with

$$\Omega := \begin{pmatrix} \Omega_{(m\times m)}^{(1)} & \Omega_{(m\times v)}^{(2)} \\ 0 & \Omega_{(v\times v)}^{(3)} \end{pmatrix}$$

for $\Omega_{(m\times m)}^{(1)}$ some regular diagonal matrix and $\Omega_{(v\times v)}^{(3)}$ some regular matrix is also a solution as $x = S\Omega^{-1}\Omega y$ holds and Ωy preserves $\widetilde{\gamma}_{i,j}^{(k)} = 0$ for $i \neq j, 1 \leq i, j \leq m$. Note that Ω^{-1} has the same form as Ω, *i.e.*

$$\Omega^{-1} = \begin{pmatrix} {\Omega_{(m\times m)}^{(1)}}^{-1} & \widetilde{\Omega}_{(m\times v)}^{(2)} \\ 0 & {\Omega_{(v\times v)}^{(3)}}^{-1} \end{pmatrix} \text{ with } \widetilde{\Omega}^{(2)} := -{\Omega^{(1)}}^{-1}\Omega^{(2)}{\Omega^{(3)}}^{-1}.$$

Thus we are able to choose Ω^{-1} such that

$$S\Omega^{-1} = \begin{pmatrix} S_{(m\times m)}^{(1)} & S_{(m\times v)}^{(2)} \\ S_{(v\times m)}^{(3)} & S_{(m\times m)}^{(4)} \end{pmatrix} \Omega^{-1} = \begin{pmatrix} \widetilde{S}_{(m\times m)}^{(1)} & 0 \\ \widetilde{S}_{(v\times m)}^{(3)} & I \end{pmatrix} \tag{6}$$

under the condition that $S^{(1)}\widetilde{\Omega}^{(2)} + S^{(2)}{\Omega^{(3)}}^{-1} = I$ and $S^{(3)}\widetilde{\Omega}^{(2)} + S^{(4)}{\Omega^{(3)}}^{-1} = 0$. Note that this is always the case because S is regular and thus $S^{(1)}||S^{(2)}$ has full rank. As ${\Omega^{(1)}}^{-1}$ is just a diagonal matrix, we are only able to fix the first row in $\widetilde{S}^{(1)}$ and thus the remaining number of free variables per column is $d_i = n - 1$ for $1 \leq i \leq m$ and $d_i = 0$ for $m + 1 \leq i \leq n$.

Corollary 1. *For $n \geq m(m+1)$ the Kipnis-Patarin-Goubin approach is upper bounded by $\mathcal{V}(B_i) = n - 1$ for $1 \leq i \leq m$ and $\mathcal{V}(B_i) = 0$ for $m + 1 \leq i \leq n$. This leads to $b_i \leq (n-2)/(i-1)$ for $1 \leq i \leq m$ and $b_i = 0$ for $m + 1 \leq i \leq n$.*

As $\mathcal{V}(B_1) \leq \ldots \leq \mathcal{V}(B_m) = m(m-1) < m(m+1) - 1 \leq n - 1$ hold, Kipnis-Patarin-Goubin were right in assuming that their system of linear equations is solvable. But as $\mathcal{V}(B_i) = 0$ for $m + 1 \leq i \leq n$ they indeed have to use (5) to eliminate the $O \times V$ coefficients.

Let us now come back to our case of figure 1. Our approach eliminates as many $O \times O$ coefficients $\widetilde{\gamma}_{i,j}^{(k)}$ with $i \neq j$ as possible using a linear transformation S of variables. After applying (5) we obtain equations that are linear in fields of characteristic two and thus can be used to substitute variables in the remaining equations. This introduces new $O \times O$ coefficients in the remaining equations and thus we skip eliminating them beforehand (see figure 3).
In order to preserve the structure of Φ, the only transformation of variables Ω applicable is mapping O variables to itself or some permutation and V variables

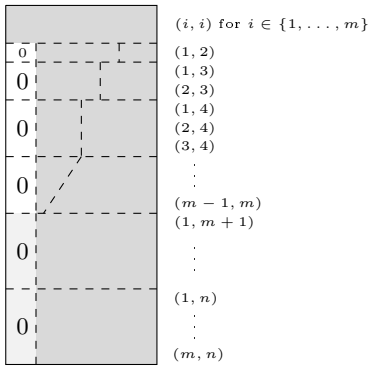

Fig. 3. Upper part of $\Phi^\intercal$ fixing only significant coefficients.

to sums of V variables. In contrast to corollary 1 we are not allowed to map O variables to V variables as this would introduce new $O \times V$ monomials due to $\widetilde{\gamma}_{i,i}^{(k)} \neq 0$. Thus Ω^{-1} is of form

$$\Omega^{-1} = \begin{pmatrix} \Omega_{(m\times m)}^{(1)^{-1}} & 0 \\ 0 & \Omega_{(v\times v)}^{(3)^{-1}} \end{pmatrix},$$

with $\Omega_{(m\times m)}^{(1)^{-1}}$ a diagonal matrix. This leads to a minimal representative of every equivalence class of the form

$$S\Omega^{-1} = \begin{pmatrix} \widetilde{S}_{(m\times m)}^{(1)} & \widetilde{S}_{(m\times v)}^{(2)} \\ \widetilde{S}_{(v\times m)}^{(3)} & I \end{pmatrix}, \tag{7}$$

where the first row of $\widetilde{S}_{(m\times m)}^{(1)}$ is fixed.

Corollary 2. *For $n = \omega m$ our approach pictured in figure 3 is upper bounded by $\mathcal{V}(B_i) = n - 1$ for $1 \leq i \leq m$ and $\mathcal{V}(B_i) = m$ for $m + 1 \leq i \leq n$. This leads to $b_i = (n-1)/(i-1)$ for $1 \leq i \leq m$ and $b_i = 1$ for $m + 1 \leq i \leq n$.*

Claim. We claim that corollary 1 as well as corollary 2 also yields a lower bound and thus are sharp. The crucial question is, if the $S\Omega^{-1}$ given by (6) or (7) is minimal for all columns i with $\mathcal{V}(B_i) = d_i$, *i.e.* we cannot find a representative $\widetilde{S}$ with less free variables in those columns we are actually using all the free variables. This is the case, if these columns are uniquely defined for some generic Π. For fixed columns 1 to $(i-1)$ this is obviously the case. Although intuitively clear, a rigorous mathematical proof seems to require stronger tools. Nevertheless, experiments prove that there are no systematic dependencies and thus corollary 1 as well as corollary 2 are tight (cf. appendix 5).

The volume of the zero blocks $B_{m+1}, \ldots, B_n$ in figure 3 is

$$mb_m = m\frac{n-1}{m-1} > n-1 > m$$

i.e. larger than the number of independent variables for $m+1 \leq i \leq n$. Thus the solution S would be singular. In order to eliminate the $O \times V$ coefficients, we also have to calculate $y_{m+1}, \ldots, y_{m+v}$ by Gaussian Elimination such that $L_{i,j} = 0$ for all $1 \leq i, j \leq m$ (see equation 5). Corollary 2 leads to $b_m = (n-1)/(m-1) > \lfloor \omega \rfloor$ for $\omega \geq 2$, *i.e.* we are able to reduce the $\mathcal{MQ}$-system to $(m - \lfloor \omega \rfloor)$ equations and variables in this case.
But in order for the $b_m m$ equations $L_{i,j} = 0$ in $(\omega - 1)m$ variables to be solvable, we have to choose $b_m = \lfloor \omega \rfloor - 1$ and thus we are only able to reduce the $\mathcal{MQ}$-system to $(m - \lfloor \omega \rfloor + 1)$ equations and variables, respectively. For $\lfloor \omega \rfloor \mid m$ our algorithm in the next section will merge both strategies, *i.e.* first eliminating some $O \times V$ coefficients and then use equation (5). This allows us to reduce to an $\mathcal{MQ}$-system with $(m - \lfloor \omega \rfloor)$ equations and variables.

4.2 Our Algorithm in the most General Form

For a very tight analysis, which give a further improvement if $\lfloor \omega \rfloor \mid m$, we first use algorithm 1 to eliminate the $O \times O$ coefficients $\widetilde{\gamma}_{i,j}^{(k)}$ with $i \neq j$ in the first $\lfloor \omega \rfloor$ equations. Corollary 2 ensures that this is possible. Next we eliminate the $O \times V$ coefficients $\widetilde{\gamma}_{i,j}^{(k)}$ for $m+1 \leq j \leq n$, $1 \leq k \leq \lfloor \omega \rfloor$ and $1 \leq i \leq (m/\lfloor \omega \rfloor)$. See figure 4 for illustration. This is possible as the number of coefficients $\widetilde{\gamma}_{i,j}^{(k)}$

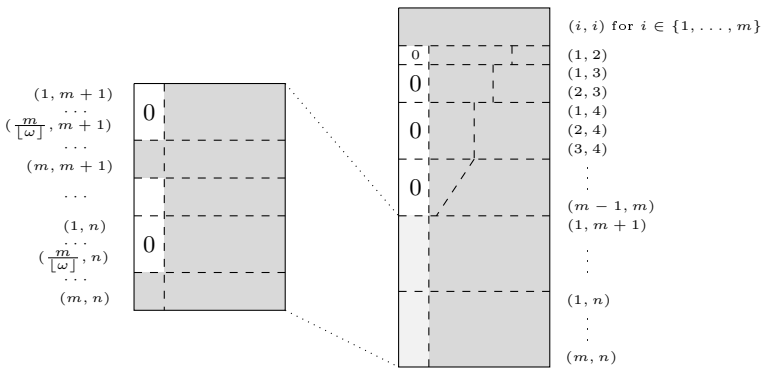

Fig. 4. Upper part of $\Phi^{\intercal}$ for tight analysis.

for $1 \leq i \leq (m/\lfloor \omega \rfloor)$ equals the number of independent variables in the j-th column of S due to $m = d_j$ (cf. corollary 2). To eliminate the linear terms $L_{i,j}$ (cf. equation (5)) we have to solve $\omega\,(m - (m/\omega)) = (\omega - 1)m$ equations, which equals the number of variables and thus yields a solution. We obtain ω equations

of the form (8). Using the Frobenius Homomorphism several times $x \mapsto x^{2^{p-1}}$ over $\mathbb{F}_{2^p}$ leads to equation (8).

$$\begin{aligned} &\sum_{i=1}^{m} \beta_{i,j} y_i^2 + c_j = 0 \qquad \text{for } j \in \{1, \ldots, \omega\} \\ \Leftrightarrow &\sum_{i=1}^{m} \beta_{i,j}^{2^{p-1}} y_i + c_j^{2^{p-1}} = 0 \end{aligned} \tag{8}$$

After using equation (8) to eliminate $\lfloor \omega \rfloor$ variables in the remaining $(m - \lfloor \omega \rfloor)$ equations we obtain a $\mathcal{MQ}$-system of $(m - \lfloor \omega \rfloor)$ variables and equations. Note that if $\frac{m}{\lfloor \omega \rfloor} \notin \mathbb{N}$ this very tight analysis fails and we are only able to eliminate $(\lfloor \omega \rfloor - 1)$ instead of $\lfloor \omega \rfloor$ variables.

5 Complexity Analysis

The complexity of our approach is on the one hand the complexity of the preprocessing step, *i.e.* applying the transformation of variables and on the other hand the complexity of solving the obtained $\mathcal{MQ}$-system by some algorithm like F_5. In the case of $\frac{m}{\lfloor \omega \rfloor} \notin \mathbb{N}$ we would have to solve $(m-1)$ systems of linear equations of different dimensions to eliminate the $O \times O$ coefficients. Deleting the $O \times V$ coefficients requires solving another linear system of size $(\lfloor \omega \rfloor - 1)m$. The overall complexity is

$$\mathcal{O}\left(\sum_{i=1}^{m} \left(\left(\lfloor \omega \rfloor - 1\right) i\right)^3\right) = \mathcal{O}\left(m(\lfloor \omega \rfloor\, m)^3\right).$$

In the tight case of $\frac{m}{\lfloor \omega \rfloor} \in \mathbb{N}$ we have to solve $(m-1)$ systems of linear equations of different dimension to eliminate the $O \times O$ coefficients and another $(\lfloor \omega \rfloor - 1)m$ systems of dimension m to delete some of the $O \times V$ coefficients. Deleting the remaining coefficients using (5) requires solving another linear system of size $(\lfloor \omega \rfloor - 1)m$. The overall complexity is

$$\mathcal{O}\left(m(\lfloor \omega \rfloor - 1)m^3 + \left((\lfloor \omega \rfloor - 1)m\right)^3 + \sum_{i=1}^{m-1} \left(\lfloor \omega \rfloor\ i\right)^3\right) = \mathcal{O}\left(m(\lfloor \omega \rfloor\, m)^3\right).$$

To determine the complexity of solving a $\mathcal{MQ}$-system using F_5 we refer to [1]. In a nutshell, we first have to calculate the degree of regularity. For semi-regular sequences, which generic systems are assumed to be, the degree of regularity is the index of the first non-positive coefficient in the Hilbert series $S_{m,n}$ with

$$S_{m,n} = \frac{\prod_{i=1}^{m}(1 - z^{d_i})}{(1-z)^n},$$

where d_i is the degree of the i-th equation. Then the complexity of solving a zero-dimensional (semi-regular) system using F_5 [2, Prop. 2.2] is

$$\mathcal{O}\left(\left(m\binom{n + d_{reg} - 1}{d_{reg}}\right)^{\alpha}\right),$$

with $2 \leq \alpha \leq 3$ the linear algebra constant. We use $\alpha = 2$ as the equations are sparse and to be comparable to the results of [2], who gave the currently best attack against UOV.

Table 1. Attack complexity against UOV, comparing the improved attack with the previously known best attack using the hybrid approach of [2], *i.e.* guessing g variables beforehand. The previously secure value for $m = 26$ is marked in **bold**. Rows where m is divisible by $\omega = 3$ are marked with "←".

	direct attack [2]			our approach					improvement	
m	g	d_{reg}	$\log_2$ complexity	$\log_2$ Gauss	g	d_{reg}	$\log_2$ complexity	overall	$\log_2$	
24	1	13	78.0	19.8	1	11	68.2	68.2	9.8	←
25	1	13	79.5	18.1	1	12	73.6	73.6	5.9	
26	1	14	**83.7**	18.3	1	13	78.1	**78.1**	5.6	
27	1	14	85.7	20.4	1	13	78.1	78.1	7.6	←
28	1	15	89.4	18.7	1	14	83.7	83.7	5.7	
29	1	15	90.6	18.9	1	14	85.1	85.1	5.5	
30	1	16	95.0	21.0	1	14	85.1	85.1	9.9	←

Table 1 give some examples of the complexity of our algorithm applied to attack UOV. The underlying field is $\mathbb{F}_{2^8}$ and nowadays parameters are $n = 78$ variables and $m = 26$ equations, *i.e.* $\omega = 3$ [2]. We use the HybridF$_5$ algorithm and thus g denote the optimal number of variables to guess. Referring to table 1, we see that today's parameter of UOV are insecure (row with **bold** values). Based on our analysis, we suggest UOV with $m = 28$ for $n = 3m$.

Table 2. Experimental complexities of our approach (TW) in seconds [s].

			$\mathbb{F}_{2^6}$		$\mathbb{F}_{2^8}$	
ω	m	n	standard	TW	standard	TW
1	3	6	0	0	0	0
1	4	8	0.01	0	0	0
1	5	10	0.03	0.02	0.03	0.02
1	6	12	0.25	0.06	0.25	0.06
1	7	14	2.94	0.33	3.10	0.34
1	8	16	33	3.51	36	3.60
1	9	18	460	43	479	45
2	3	9	0	0	0	0
2	4	12	0.01	0	0	0.03
2	5	15	0.05	0.07	0.05	0.08
2	6	18	0.35	0.29	0.36	0.28
2	7	21	3.40	0.80	3.45	0.80
2	8	24	38	2.19	41	2.25
2	9	27	520	8.34	546	8.10

We have implemented our algorithm using the software system Magma V2.16-1 [4] and found it to be in line with the theoretical predictions. All experiments were performed on a Intel Xeon X33502.66GHz (Quadcore) with 8 GB of RAM using only one core. Table 2 compares the time complexities of the standard approach of guessing v variables and solve the remaining $\mathcal{MQ}$-system in m variables and our algorithm for various parameter sets. The source code of our implementation can be found on the homepage of the first author.

6 Odd Cases

In this section we outline some ideas to extend our results to fields of *odd* characteristic. Hence we are now working over $\mathbb{F}_{p^k}$ for some prime $p \neq 2$ and $k \in \mathbb{N}_{>0}$. Unfortunately there is no straightforward extension of our ideas. The main problem is that equations (8) are not longer linear and thus we are not able to eliminate variables in the remaining equations. Nevertheless, Gröbner algorithms are empirically faster on systems containing equations (8), but it is hard to quantify the gain from a theoretical perspective. An argument that this task is inherently difficult is also the odd-characteristic algorithm of Courtois *et al.* [6]. It extended the algorithm for even characteristics by Kipnis-Patarin-Goubin [15] to the odd case. However, it requires now $n \geq 2^{\frac{m}{7}}(m+1)$—which is infeasible in practice. However, for *small* values of ω, we can actually adapt our algorithm from even to odd characteristics. This coincides with the cryptanalytically interesting case of UOV, where we have $\omega \approx 3$ for efficiency reasons. Our main concern is to obtain some equations $y_i = g(y_{\omega+1}, \ldots, y_m)$ for $1 \leq i \leq \omega$ and some polynomial function g of low degree from the ω equations given by (8). They will be used to eliminate the variables $y_1, \ldots, y_\omega$. Therefore, we need to determine coefficients $(\beta_{\omega+1,i}, \ldots, \beta_{m,i})$ such that they are linearly dependent on $(\beta_{\omega+1,1}, \ldots, \beta_{m,1})$ for every $i \in \{2, \ldots, \omega\}$. This way we could eliminate these parts in equations 2 to ω by Gaussian Elimination. Producing an upper triangular form on these equations leads to $y_i^2 = c_i$ for $2 \leq i \leq \omega$, which is efficiently solvable for finite fields of size p^k. Still, the question remains how to determine the coefficients β_{ij}. Fixing $\widetilde{\gamma}_{i,i}^{(j)} = \beta_{ij}$ to some value for $2 \leq j \leq \omega$ leads to a *quadratic* system of $(\omega - 1)$ equations and variables s_{ij} (cf. figure 1)—so we seem to be back on square one. However, if ω is sufficiently small, *i.e.* smaller than 20, we can use any $\mathcal{MQ}$-solver, such as Gröbner algorithms for this task.

7 Conclusions and Open Questions

In this article we showed a more "gradual" change between exponential running time in the determined case ($n = m$) and the polynomial running time in massively underdetermined case ($n \geq m(m+1)$). Previously, this change was abrupt (Kipnis-Goubin-Patarin), *i.e.* there was a polynomial time algorithm in one case, and a fully exponential algorithm in the other. The situation is depicted in figure 5.

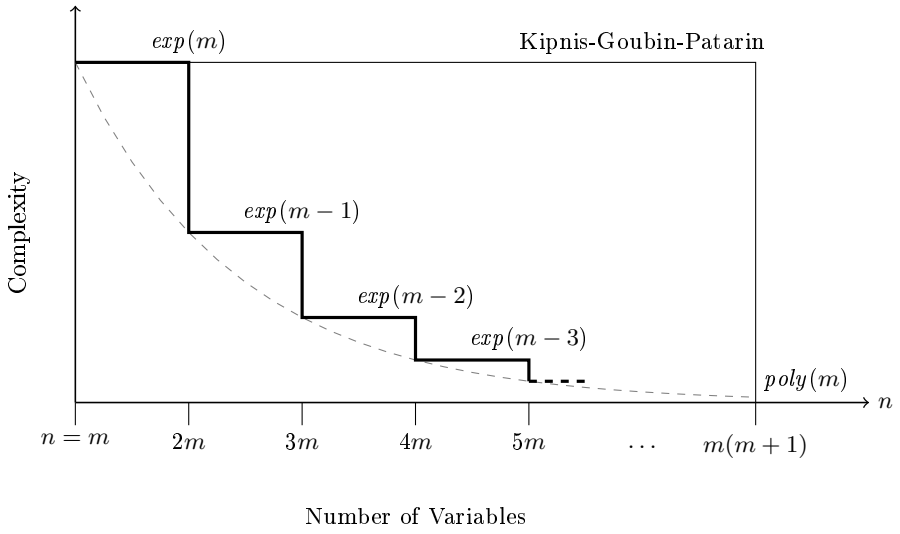

Fig. 5. Achievement of this paper (solid), compared with Kipnis-Goubin-Patarin (thin) for fixed m.

Our algorithm can be used as a general preprocessing step for further applications. Applied to UOV we would have to raise parameters from $m = 26$ to $m = 28$ in order to make the scheme secure.

Acknowledgments

We want to thank Gottfried Herold (Bochum) for fruitful discussions and helpful remarks. Furthermore we thank the reviewers for helpful comments.
The authors were supported by the German Science Foundation (DFG) through an Emmy Noether grant where the second author is principal investigator. All authors were in part supported by the European Commission through the IST Programme under contract *ICT-2007-216676 Ecrypt II*.

Bibliography

[1] M. Bardet, J.-C. Faugère, B. Salvy, and B.-Y. Yang. Asymptotic behaviour of the degree of regularity of semi-regular polynomial systems. In P. Gianni, editor, *MEGA 2005 Sardinia (Italy)*, 2005.

[2] L. Bettale, J.-C. Faugère, and L. Perret. Hybrid approach for solving multivariate systems over finite fields. *In Journal of Mathematical Cryptology*, 3:177–197, 2009.

[3] A. Braeken, C. Wolf, and B. Preneel. A study of the security of Unbalanced Oil and Vinegar signature schemes. In *The Cryptographer's Track at RSA Conference 2005*, volume 3376 of *Lecture Notes in Computer Science*. Alfred J. Menezes, editor, Springer, 2005. 13 pages, cf `http://eprint.iacr.org/2004/222/`.

[4] Computational Algebra Group, University of Sydney. *The MAGMA Computational Algebra System for Algebra, Number Theory and Geometry*. `http://magma.maths.usyd.edu.au/magma/`.

[5] N. Courtois. Higher order correlation attacks, XL algorithm and cryptanalysis of Toyocrypt. In *ICISC*, volume 2587 of *Lecture Notes in Computer Science*, pages 182–199. Pil Joong Lee and Chae Hoon Lim, editors, Springer, 2002.

[6] N. Courtois, L. Goubin, W. Meier, and J.-D. Tacier. Solving underdefined systems of multivariate quadratic equations. In *Public Key Cryptography — PKC 2002*, volume 2274 of *Lecture Notes in Computer Science*, pages 211–227. David Naccache and Pascal Paillier, editors, Springer, 2002.

[7] N. T. Courtois, M. Daum, and P. Felke. On the security of HFE, HFEv- and Quartz. In *Public Key Cryptography — PKC 2003*, volume 2567 of *Lecture Notes in Computer Science*, pages 337–350. Y. Desmedt, editor, Springer, 2002. `http://eprint.iacr.org/2002/138`.

[8] N. T. Courtois, A. Klimov, J. Patarin, and A. Shamir. Efficient algorithms for solving overdefined systems of multivariate polynomial equations. In *Advances in Cryptology — EUROCRYPT 2000*, volume 1807 of *Lecture Notes in Computer Science*, pages 392–407. Bart Preneel, editor, Springer, 2000. Extended Version: `http://www.minrank.org/xlfull.pdf`.

[9] J.-C. Faugère. A new efficient algorithm for computing Gröbner bases (F_4). *Journal of Pure anNational Institute of Standards and Technologyd Applied Algebra*, 139:61–88, June 1999.

[10] J.-C. Faugère. A new efficient algorithm for computing Gröbner bases without reduction to zero (F_5). In *International Symposium on Symbolic and Algebraic Computation — ISSAC 2002*, pages 75–83. ACM Press, July 2002.

[11] J.-C. Faugère and A. Joux. Algebraic cryptanalysis of Hidden Field Equations (HFE) using Gröbner bases. In *Advances in Cryptology — CRYPTO 2003*, volume 2729 of *Lecture Notes in Computer Science*, pages 44–60. Dan Boneh, editor, Springer, 2003.

[12] J.-C. Faugère, A. Otmani, L. Perret, and J.-P. Tillich. Algebraic cryptanalysis of McEliece variants with compact keys. In *EUROCRYPT*, volume 6110 of *Lecture Notes in Computer Science*, pages 279–298. Henri Gilbert, editor, Springer, 2010. ISBN 978-3-642-13189-9.

[13] S. Fischer and W. Meier. Algebraic immunity of S-boxes and augmented functions. In *FSE*, volume 4593 of *Lecture Notes in Computer Science*, pages 366–381. Alex Biryukov, editor, Springer, 2007. ISBN 978-3-540-74617-1.

[14] M. R. Garey and D. S. Johnson. *Computers and Intractability — A Guide to the Theory of NP-Completeness*. W.H. Freeman and Company, 1979. ISBN 0-7167-1044-7 or 0-7167-1045-5.

[15] A. Kipnis, J. Patarin, and L. Goubin. Unbalanced Oil and Vinegar signature schemes — extended version, 2003. 17 pages, `citeseer/231623.html`, 2003-06-11.

[16] S. Murphy and M. J. Robshaw. Essential algebraic structure within the AES. In *Advances in Cryptology — CRYPTO 2002*, volume 2442 of *Lecture Notes in Computer Science*, pages 1–16. Moti Yung, editor, Springer, 2002.

[17] J. Patarin. The oil and vinegar signature scheme. Presented at the Dagstuhl Workshop on Cryptography, Sept. 1997. transparencies.

[18] A. Petzoldt, E. Thomae, S. Bulygin, and C. Wolf. Small public keys and fast verification for multivariate quadratic public key systems. In *Proceedings of Workshop on Cryptographic Hardware and Embedded Systems 2011 (CHES 2011)*, Lecture Notes in Computer Science. Springer, Sep 2011.

[19] M. Sugita, M. Kawazoe, L. Perret, and H. Imai. Algebraic cryptanalysis of 58-round sha-1. In *FSE*, pages 349–365, 2007.

[20] C. Wolf and B. Preneel. Equivalent keys in HFE, C^*, and variations. In *Proceedings of Mycrypt 2005*, volume 3715 of *Lecture Notes in Computer Science*, pages 33–49. Serge Vaudenay, editor, Springer, 2005. Extended version `http://eprint.iacr.org/2004/360/`, 15 pages.

[21] C. Wolf and B. Preneel. Equivalent keys in multivariate quadratic public key systems. *Journal of Mathematical Cryptology*, 4(4):375–415, 2011.