# Hybrid Approach : a Tool for Multivariate Cryptography

Luk Bettale*, Jean-Charles Faugère and Ludovic Perret

INRIA, Centre Paris-Rocquencourt, SALSA Project
UPMC, Univ. Paris 06, LIP6
CNRS, UMR 7606, LIP6
Boîte courrier 169
4, place Jussieu
75252 Paris Cedex 05, France
luk.bettale@lip6.fr

**Abstract.** In this paper, we present an algorithmic tool to cryptanalysis multivariate cryptosystems. The presented algorithm is a hybrid approach that mixes exhaustive search with classical Gröbner bases computation to solve multivariate polynomial systems over a finite field. Depending on the size of the field, our method is an improvement on existing techniques. For usual parameters of multivariate schemes, our method is effective. We give theoretical evidences on the efficiency of our approach as well as practical cryptanalysis of several multivariate signature schemes (TRMS, UOV) that were considered to be secure. For instance, on TRMS, our approach allow to forge a valid signature in $2^{67}$ operations instead of $2^{160}$ with exhaustive search or $2^{83}$ with only Gröbner bases. Our algorithm is general as its efficiency is demonstrated on random systems of equations. As the structure of the cryptosystem is not involved, our algorithm provides a generic tool to calibrate the parameters of any multivariate scheme. These results were already published in [5]. We also present an extended version of our hybrid approach, suitable for polynomials of higher degree. To easily access our tools, we provide a MAGMA package available at `http://www-salsa.lip6.fr/~bettale/hybrid.html` that provide all the necessary material to use our hybrid approach and to compute the complexities.

## 1 Introduction

Multivariate cryptography is a family of public key cryptosystems. The idea is to present the public key as a set of (generally quadratic) polynomials in a large number of variables. To introduce a trapdoor, a special algebraic system $F = (f_1(x_1,\ldots,x_n),\ldots,f_m(x_1,\ldots,x_n))$ is built such that it is easy to invert. The classical trapdoors are STS, UOV or HFE. To hide the structure of $F$, two invertible affine transformations $S,T \in \mathrm{Aff_n}(\mathbb{K})$ are chosen and the public key is the system

$$G = g_1(x_1,\ldots,x_n),\ldots,g_m(x_1,\ldots,x_n) = T \circ F \circ S.$$

To encrypt, the system $G$ is evaluated in the variables $m_1,\ldots,m_n$ corresponding to a message. The knowledge of the private key $F,S,T$ allows the legitimate recipient to efficiently recover the message whereas an attacker has to solve the algebraic system $G$ which should have no visible structure.

The problem of solving a multivariate system of equations, a.k.a. PoSSo, is known to be NP-hard, and also hard in average (exponential time). Note that PoSSo remains NP-hard even if the input polynomials are quadratics. In this case, PoSSo is also called $\mathscr{MQ}$. The security of a multivariate scheme relies directly on the hardness of solving a multivariate algebraic system of equations. In this context, it is important to have efficient tools to solve polynomial systems. When the system is considered as hard to solve as a random one (which is ideally required for a multivariate system), only general tools can be used to solve the system. We present in this paper an improved tool, namely the hybrid approach, that does not take advantage on the structure of the equations, but rather of the context to enhance the polynomial system solving. We use the fact that the field of coefficient is finite to perform a mix of exhaustive search and classical Gröbner bases techniques. For the parameters used in cryptography, our analysis shows that the hybrid approach brings a significant improvement over the classical methods. In [7], the authors did not

* author partially supported by DGA/MRIS (french secretary of defense)

succeed to attack UOV with Gröbner bases. Indeed, the parameters were unreachable using a standard zero-dimensional approach. With the hybrid approach we were able to break these parameters. Using this algorithm, we can put the security of general multivariate schemes to the proof. As our theoretic analysis allows to refine the security parameters, this make it a useful tool to design or cryptanalyze multivariate schemes.

Not only multivariate cryptography is concerned by the hybrid approach. In [15], the authors give a new method to solve the discrete logarithm problem on the group of points of an elliptic curve defined over an extension field. To do so, they have to solve a system of equations with of high degree in a quite big field. We present in this paper an extended hybrid approach which could be more suitable for this kind of problems.

Our contributions are available at `http://www-salsa.lip6.fr/~bettale/hybrid.html`

**Organization of the paper**

The paper is organized as follows. After this introduction, we present the general problem of solving a polynomial system as well as the classical method to address it, namely the zero-dim solving strategy using Gröbner bases. We also give the definitions of semi-regular sequences and degree of regularity, necessary to compute the complexity of our approach. In Section 3, we present the hybrid approach algorithm as well as its complexity. In Section 4, we give a generalization of the hybrid approach that uses splitted field equations. The scope of the extended hybrid approach will not be the same as the classical hybrid approach as it will be more efficient on polynomial systems of higher degree.

## 2 Polynomial System Solving

The general problem is to find (if any) $(z_1,\ldots,z_n) \in \mathbb{K}^n$ such that:

$$\begin{cases} f_1(z_1,\ldots,z_n) = 0 \\ \quad\vdots \\ f_m(z_1,\ldots,z_n) = 0 \end{cases}$$

The best known method is to compute the Gröbner basis of the ideal generated by this system. We refer the reader to [1, 10] for a more thorough introduction to ideals and Gröbner bases. Informally, a Gröbner basis is a set of generators of an ideal which has "good" properties. In particular, if the system has a finite number of solution (zero-dimensional ideal), a Gröbner basis in Lex order has the following shape:

$$\{g_1(x_1),\ldots,g_2(x_1,x_2),\ldots,g_{k_1}(x_1,x_2),g_{k_1+1}(x_1,x_2,x_3),\ldots,g_{k_n}(x_1,\ldots,x_n)\}.$$

With this special structure, the system may be easily solved by successively eliminating variables, namely computing solutions of univariate polynomials and back-substituting the results.

The historical method for computing Gröbner bases was introduced by Buchberger in [8, 9]. Many improvements has been done leading to more efficient algorithms such as $F_4$ and $F_5$ due to Faugère [11, 12]. The algorithm $F_4$ for example is the default algorithm for computing Gröbner bases in the computer algebra softwares MAGMA and MAPLE. The $F_5$ algorithm[1] is even more efficient. We have mainly used this algorithm in our experiments. For our purpose, it is not necessary to describe the algorithm, but we give its complexity.

**Proposition 1.** *The complexity of computing a Gröbner basis of a zero-dimensional system of m equations in n variables with $F_5$ is:*

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

*where $d_{reg}$ is the degree of regularity of the system and $2 \leq \omega \leq 3$ is the linear algebra constant.*

[1] available through FGb

From a practical point of view, it is much faster to compute a Gröbner basis for a degree ordering such as the Degree Reverse Lexicographic (DRL) order than for a Lexicographic order (Lex). For zero-dimensional systems, it is usually less costly to first compute a DRL-Gröbner basis, and then to compute the Lex-Gröbner basis using a change ordering algorithm such as FGLM [13]. This strategy called zero-dim solving is performed blindly in modern computer algebra softwares. This is convenient for the user, but can be an issue for advanced users.

**Proposition 2.** *Given a Gröbner basis $G_1 \subset \mathbb{K}[x_1,\ldots,x_n]$ w.r.t. a monomial ordering $\prec_1$ of a zero-dimensional system the complexity of computing a Gröbner basis $G_2 \subset \mathbb{K}[x_1,\ldots,x_n]$ w.r.t. a monomial ordering $\prec_2$ with FGLM is:*

$$\mathcal{O}\left(n \cdot D^{\omega}\right)$$

*where $D$ is the degree of the ideal generated by $G_1$ (i.e. the number of solutions counted with multiplicity in the algebraic closure of $\mathbb{K}$).*

We see easily that the cost of change ordering is negligible when the system has very few solutions.

For a finite field $\mathbb{K}$ with $q$ elements, one can always add the field equations $x_1^q - x_1, \ldots, x_n^q - x_n$ to explicitly look for solutions over the ground field $\mathbb{K}$ and not in some extensions. By doing this, we will always obtain an over-defined system. This technique is widely used, and improves the computation of solutions if $q \ll n$. Otherwise, the addition of the field equations does not lead to a faster computation of a Gröbner basis. Even worse, this can slow down the computation due to the high degrees of the equations. In multivariate cryptography, some schemes use for example the field $\mathbb{F}_{2^8}$ whose elements can easily be represented with a byte. The hybrid method that we will present is especially suitable in such situation.

### 2.1 Semi-regular sequences

In order to study random systems, we need to formalize the definition of "random systems". To do so, the notion of regular sequences and semi-regular sequences (for over-defined systems) has been introduced in [2]. We give the definition here.

**Definition 1.** *Let $\{p_1,\ldots,p_m\} \subset \mathbb{K}[x_1,\ldots,x_n]$ be homogeneous polynomials of degrees $d_1,\ldots,d_m$ respectively. This sequence is semi-regular if:*

- $\langle p_1,\ldots,p_m\rangle \neq \mathbb{K}[x_1,\ldots,x_n]$
- *for all $1 \leq i \leq m$ and $g \in \mathbb{K}[x_1,\ldots,x_n]$:*

$$\deg(g \cdot p_i) < d_{reg} \text{ and } g \cdot p_i \in \langle p_1,\ldots,p_{i-1}\rangle \Rightarrow g \in \langle p_1,\ldots,p_{i-1}\rangle.$$

This notion can be extended to affine polynomials by considering their homogeneous components of highest degree. It has been proven in [2, 3] that for semi-regular sequences, the degree of regularity can be computed explicitly.

*Property 1.* The degree of regularity of a semi-regular sequence $p_1,\ldots,p_m$ of respective degrees $d_1,\ldots,d_m$ is given by the index of the first non-positive coefficient of:

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

Let $D = \{d_1,\ldots,d_m\}$, we will denote the degree of regularity by $\mathrm{d_{reg}}\left(n,m,D\right)$.

This property allows us to have a very precise knowledge of the complexity of the computation of a Gröbner basis for semi-regular systems. For semi-regular systems it has been proven that the degree decreases as $m$ goes larger. Thus, the more a system is over-defined, the faster its Gröbner basis can be computed.

For more convenience, we denote from now on the complexity of $\mathrm{F_5}$ for semi-regular systems of equations of degree $d_1,\ldots,d_m$ as the function

$$\mathrm{C_{F_5}}\left(n,m,D\right) = \left(m \cdot \binom{n+\mathrm{d_{reg}}(n,D)-1}{\mathrm{d_{reg}}(n,D)}\right)^{\omega}$$

where $D$ is the set $\{d_1,\ldots,d_m\}$.

## 3 Hybrid Approach

In many cases (especially in multivariate cryptography), the coefficient field is much bigger than the number of variables. In this case, as we have seen in Section 2, adding the field equations can dramatically slow down the computation of a Gröbner basis.

We present in this section our hybrid approach mixing exhaustive search and Gröbner bases techniques. First we will present the algorithm and discuss its complexity. Its efficiency depends on the choice of a proper trade-off. We take advantage of the behavior of semi-regular systems to find the best trade-off. After that, we give some examples coming from proposed cryptosystems as proof of concept.

### 3.1 Algorithm

In a finite field, one can always find all the solutions of an algebraic system by exhaustive search. The complete search should take $q^n$ evaluations of the system if $n$ is the number of variables and $q$ the size of the field. The idea of the hybrid approach is to mix exhaustive search with Gröbner basis computations. Instead of computing one single Gröbner basis of the whole system, we compute the Gröbner bases of $q^k$ subsystems obtained by fixing $k$ variables. The intuition is that the gain obtained by solving systems with less variables may overcome the loss due to the exhaustive search on the fixed variables. Algorithm 1 describes the hybrid approach.

**Algorithm 1** HybridSolving

**Input:** $\mathbb{K}$ is finite, $\{f_1,\ldots,f_m\} \subset \mathbb{K}[x_1,\ldots,x_n]$ is zero-dimensional, $k \in \mathbb{N}$.
**Output:** $\mathscr{S} = \{(z_1,\ldots,z_n) \in \mathbb{K}^n : f_i(z_1,\ldots,z_n) = 0, 1 \leq i \leq m\}$.

```
S := ∅
for all (v_1,...,v_k) ∈ K^k do
    Find the set of solutions S' ⊂ K^(n-k) of
    f_1(x_1,...,x_{n-k},v_1,...,v_k) = 0,..., f_m(x_1,...,x_{n-k},v_1,...,v_k) = 0
    using the zero-dim solving strategy.
    S := S ∪ {(z'_1,...,z'_{n-k},v_1,...,v_k) : (z'_1,...,z'_{n-k}) ∈ S'}.
end for
return S.
```

As for the $F_5$ algorithm, the complexity of Algorithm 1 can be determined if the system is semi-regular. However, as the algorithm deals with sub-systems of $m$ equations in $n-k$ variables, we will make the following assumption.

**Hypothesis 1** *Let $\mathbb{K}$ be a finite field and $\{f_1,\ldots,f_m\} \subset \mathbb{K}[x_1,\ldots,x_n]$ a generic semi-regular system of equations of degree $d$. We will suppose that the systems*

$$\left\{\{f_1(x_1,\ldots,x_{n-k},v_1,\ldots,v_k),\ldots,f_m(x_1,\ldots,x_{n-k},v_1,\ldots,v_k)\} : (v_1,\ldots,v_k) \in \mathbb{K}^k\right\}$$

*are semi-regular, for all $0 \leq k \leq n$.*

This hypothesis is consistent with the intuition that when some variables of a random system are fixed, the system is still random. This hypothesis has been verified with a rather large amount of random systems as well as systems coming from the applications of Section 3.2. In practice, the constructed systems may even be easier to solve than a semi-regular system. We have observed that its degree of regularity is always lower than a random system. Thus, our hypothesis can be used as it provides an upper bound on the complexity of our approach.

**Proposition 3.** *Let $\mathbb{K}$ be a finite field and $\{f_1,\ldots,f_m\} \subset \mathbb{K}[x_1,\ldots,x_n]$ be a semi-regular system of equations of degree $d_1,\ldots,d_m$ and $0 \leqslant k \leqslant n$. The complexity of solving the system with a hybrid approach, is bounded from above by:*

$$\mathscr{O}\left((\#\mathbb{K})^k \cdot \mathrm{C}_{\mathrm{F}_5}(n-k,m,D)\right)$$

*where $D = \{d_1,\ldots,d_m\}$.*

There exists a value $k$ such that the complexity from Proposition 3 is minimal. If this value is non-trivial ($k \neq 0$ and $k \neq n$) then our method is an improvement on known techniques. In the next subsection, we give theoretical evidences that our approach is relevant on some ranges of parameters. In [5], we give for quadratic systems an asymptotic analysis of this complexity and an approximation of the best trade-off with respect to the parameters. We also show that our approach brings an improvement for quadratic systems if $\log_2(q)$ is smaller than $0.6226 \cdot \omega \cdot n$ where $q$ is the size of the field and $\omega$ the linear algebra constant. For instance, to solve a system of 20 quadratic equations in 20 variables, the hybrid approach will bring an improvement if the field has a size below $2^{24}$. Theses kind of parameters are generally found in multivariate cryptography. We show in the next section how the hybrid approach permits to break the parameters of some cryptosystems.

### 3.2 Applications

As proof of concept, we applied our hybrid approach to several multivariate cryptosystems. This permits to show a weakness in the choice of the parameters the TRMS [4] and UOV [14]. Our results have been given in [5]. In this paper, we don't describe the cryptosystems and only give a summary. As our approach does not depend on the structure of the systems, only the set of parameters matters to compute upper bounds. We give in Figure 1 the complexity of our approach depending on the parameter $k$. We see the best trade-off is to choose $k = 1$. The theoretical complexity drops from $2^{80}$ to $2^{67}$. In practice, for TRMS we have even better results (reported in Table 1). In practice, choosing the best theoretical trade-off $k = 1$ would have taken too much memory, only $k = 2$ has been achieved.

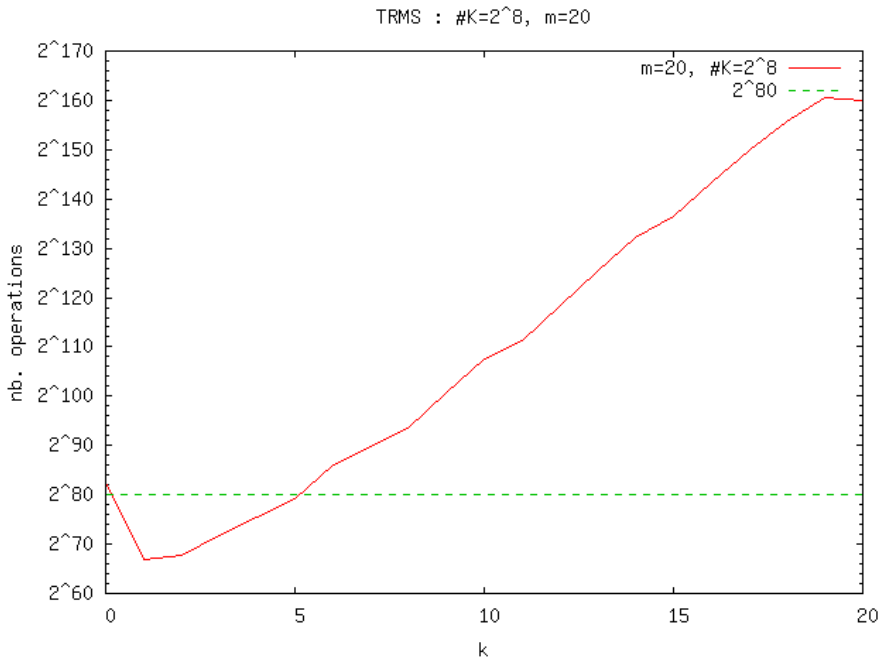


**Fig. 1.** TRMS: Complexity of hybrid approach depending on $k$

Finally, our work permits to analyze the security of several multivariate schemes only by looking at their parameters. For example, in [6], the authors proposed implementations of some multivariate schemes. We were able to compute the minimum complexity of solving the public systems and we show that for a suitable value of $k$, the complexity of breaking all the proposed parameters are below $2^{80}$. Our approach can be viewed as a tool to calibrate the parameters of multivariate cryptosystems.

**Table 1.** Experimental results on TRMS. The column $m$ is the number of variables (and equations), $m-k$ is the number of variables left after fixing $k$ variables. The columns $T_{F_5}$, $Mem_{F_5}$, and $Nop_{F_5}$ are respectively the time, memory and number of operations needed to compute one Gröbner basis with the $F_5$ algorithm. The value $T_{F_5}$ has to be multiplied by $q^k$ to obtain $Nop$, the total number of operations of the hybrid approach.

| $m$ | $m-k$ | $q^k$ | $T_{F_5}$ | $Mem_{F_5}$ | $Nop_{F_5}$ | $Nop$ |
|---|---|---|---|---|---|---|
| 20 | 18 | $2^{16}$ | 51h | 41.940 GB | $2^{41}$ | $2^{57}$ |
| 20 | 17 | $2^{24}$ | 2h45min | 4.402 GB | $2^{37}$ | $2^{61}$ |
| 20 | 16 | $2^{32}$ | 626 s. | 912 MB | $2^{34}$ | $2^{66}$ |
| 20 | 15 | $2^{40}$ | 46 s. | 368 MB | $2^{30}$ | $2^{70}$ |

# 4 Extended Hybrid Approach

In this section, we present a generalization of the hybrid approach.

## 4.1 Algorithm

We recall that the basic approach to find the solutions lying in the coefficient field $\mathbb{F}_q$ of a system of equations $f_1,\ldots,f_m$ is to solve the system with the field equations $x_1^q-x_1=0,\ldots,x_n^q-x_n=0$. When $q$ is too big, adding the field equations can be an issue.

The basis of the hybrid approach presented in Section 3 is to solve a set of easier systems of equations by fixing $k$ variables $x_1,\ldots,x_k$ to some values $v_1,\ldots,v_k$. From another point of view, this means solving the original system on which we add $k$ linear equations $x_1-v_1=0,\ldots,x_k-v_k=0$.

An idea in between could be to add "split" field equations. For a field $\mathbb{K}$ with $q$ elements, it holds that $\prod_{e\in\mathbb{K}} x-e = x^q-x$. For a given parameter $d$, one could add only parts of the field equations $\prod_{i=1}^{i\leqslant d} x-e_i$ with $e_1,\ldots,e_d\in\mathbb{K}^n$. As in the hybrid approach, we could only add $k$ equations to avoid a too big exhaustive search. The extended hybrid approach thus has two parameters. The number of split equations to be added $0\leqslant k\leqslant n$ and their maximum degree $1\leqslant d\leqslant q$. We remark that when $k=0$, it becomes the classical zero-dim solving approach, when $k=n$ and $d=q$, it is the field equations approach and when $d=1$, the approach is similar to the hybrid approach. Algorithm 2 describes the extended hybrid approach.

**Algorithm 2** ExtHybridSolving

**Input:** $\{f_1,\ldots,f_m\}\subset\mathbb{K}[x_1,\ldots,x_n]$ (zero-dim), $k,d\in\mathbb{N}$.
**Output:** $\mathscr{S}=\{(z_1,\ldots,z_n)\in\mathbb{K}^n : f_i(z_1,\ldots,z_n)=0, 1\leq i\leq m\}$.

```
S := ∅.
Let L = {h_1,...,h_l} a factorization of the field equation x^q − x with deg(h_i) ⩽ d
for all (h_{i_1},...,h_{i_k}) ∈ L^k do
    Find the set of solutions S' ⊂ K^n of
    f_1 = 0,...,f_m = 0, h_{i_1}(x_1) = 0,...,h_{i_k}(x_k) = 0
    using the zero-dim solving strategy.
    S := S ∪ S'.
end for
return S.
```

The complexity of Algorithm 2 can be computed in a similar way as Proposition 3.

**Proposition 4.** *Let $\mathbb{K}$ be a finite field and $\{f_1,\ldots,f_m\}\subset\mathbb{K}[x_1,\ldots,x_n]$ be a semi-regular system of equations of degree $d_1,\ldots,d_m$. The complexity of solving the system with the extended hybrid approach, is bounded*

*from above by:*

$$\mathscr{O}\left(\sum_{i=0}^{k}\binom{k}{i} l^{k-i}\,\mathrm{C}_{\mathrm{F}_5}\left(n, \{d_1,\ldots,d_m,\underbrace{d,\ldots,d}_{k-i},\underbrace{r,\ldots,r}_{i}\}\right)\right)$$

*where* $q = d\cdot l + r$, $0 < r \leqslant d$.

*Proof.* A field equation $x_i^q - x_i$ is split into $l$ equations of degree $d$ and 1 equation of degree $r$. For each subsystem, $i$ (over $k$) split field equations of degree $r$ are fixed, there are $l^{k-i}$ possible systems. As there are $\binom{k}{i}$ possible positions for the degree $r$ split field equations, we obtain the above result.

The above complexity can be bounded again by

$$\mathscr{O}\left(\left\lceil\frac{q}{d}\right\rceil^{k}\cdot\mathrm{C}_{\mathrm{F}_5}\left(n, \{d_1,\ldots,d_m,\underbrace{d,\ldots,d}_{k}\}\right)\right)$$

The two values match when $d \mid q$.

Here again, it is clear that the efficiency of this approach depends on the choice of parameters $k$ and $d$. In the next section we will analyze the behavior of this approach and find out how to choose proper parameters that will bring the best trade-off between exhaustive search and Gröbner bases.

### 4.2 Analysis

To analyze the behavior of our approach, we have to be able to compute exactly the complexity of solving a given system. For semi-regular systems, we can know in advance its degree of regularity, and thus the complexity of the Gröbner basis computation. To perform our analysis, we use the approximation of the degree of regularity of an over-defined system ($n$ variables, $n+k$ equations) given in [2]:

$$d_{reg} = \sum_{i=1}^{n+k}\frac{d_i-1}{2} - \alpha_k\sqrt{\sum_{i=1}^{n+k}\frac{d_i^2-1}{6}} + \mathscr{O}(1)$$

when $n\to\infty$. Here, $\alpha_k$ is the largest root of the $k$-th Hermite's polynomial. To simplify the analysis, we will use the upper bound on the complexity of the extended hybrid approach.

$$C_{Hyb} = \left\lceil\frac{q}{d}\right\rceil^{k}\cdot\mathrm{C}_{\mathrm{F}_5}\left(n, \{d_1,\ldots,d_m,\underbrace{d,\ldots,d}_{k}\}\right).$$

Using the Stirling approximation $n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$, we can compute the logarithmic derivative of $C_{Hyb}$ and thus find the minimum of the function, in the same way as in [5] for the hybrid approach.

The scope of the extended hybrid approach is not the same as the basic hybrid approach. While the hybrid approach was suitable for quadratic systems, the extended hybrid approach will show an improvement for system of equations of higher degree. For example, for a system of 5 equations of degree 8 in 5 variables in $\mathbb{F}_{31}$, the best theoretical trade-off is to add one split equation of degree 5 ($k=1$, $d=5$). From our experiments, for quadratic systems, the basic hybrid approach will always be better.

## 5 Conclusion

In this paper, we present a general tool to solve polynomial systems over finite fields, namely the hybrid approach. We have computed explicitly the complexity of this approach. The relevancy of our method is theoretically supported by the asymptotic analysis given in [5]. In practice, our approach is also efficient, in particular, it permits to break the parameters of several multivariate cryptosystems. We also present a generalization of this approach called the extended hybrid approach. From our analysis, this extension does

not overpass the hybrid approach for quadratic systems. However, the extended hybrid approach is relevant on equations of higher degree. Finally, this paper gives a toolbox to analyze the parameters of multivariate cryptosystems. The complexity of our approaches can be used to better calibrate the parameters of multivariate cryptosystems. An implementation of the hybrid approach as well as functions to easily compute the complexity of our approach are available at `http://www-salsa.lip6.fr/~bettale/hybrid.html`.

## References