Inverting HFE is Quasipolynomial

Louis Granboulan[1], Antoine Joux[2,3], and Jacques Stern[1] **

[1] École normale supérieure
Département d'Informatique 45, rue d'Ulm
75230 Paris Cedex 05, France
Louis.Granboulan@ens.fr, Jacques.Stern@ens.fr
[2] DGA
[3] Université de Versailles St-Quentin-en-Yvelines
PRISM
45, avenue des Etats-Unis
78035 Versailles Cedex, France
Antoine.Joux@m4x.org

Abstract. In the last ten years, multivariate cryptography has emerged as a possible alternative to public key cryptosystems based on hard computational problems from number theory. Notably, the HFE scheme [17] appears to combine efficiency and resistance to attacks, as expected from any public key scheme. However, its security is not yet completely understood. On one hand, since the security is related to the hardness of solving quadratic systems of multivariate binary equations, an NP complete problem, there were hopes that the system could be immune to subexponential attacks. On the other hand, several lines of attacks have been explored, based on so-called relinearization techniques [12, 5], or on the use of Gröbner basis algorithms [7]. The latter approach was used to break the first HFE Challenge 1 in 96 hours on a 833 MHz Alpha workstation with 4 Gbytes of memory. At a more abstract level, Faugère and Joux discovered an algebraic invariant that explains why the computation finishes earlier than expected. In the present paper, we pursue this line and study the asymptotic behavior of these Gröbner basis based attacks. More precisely, we consider the complexity of the decryption attack which uses Gröbner bases to recover the plaintext and the complexity of a related distinguisher. We show that the decryption attack has a quasipolynomial complexity, where quasipolynomial denotes an subexponential expression much smaller than the classical subexponential expressions encountered in factoring or discrete logarithm computations. The same analysis shows that the related distinguisher has provable quasipolynomial complexity.

1 Introduction

In the last ten years, multivariate cryptography has emerged as a possible alternative to public key cryptosystems based on hard computational

** This work is supported in part by the French government through X-Crypt, in part by the European Commission through ECRYPT.

problems from number theory. The public key of multivariate schemes is a system of multivariate quadratic (MQ) equations over a finite field, and the underlying question of finding a solution to such systems, a well known NP-complete problem, seems to form a basis for the security of the schemes, the same way RSA-like cryptosystems have their security based on the hardness of factoring, and ElGamal-like cryptosystems rely on the discrete logarithm problem.

Although all MQ schemes are more or less built on the same pattern, mixing the equations and the unknowns coming from a trapdoor *internal MQ function* using invertible affine transforms, there exist many such schemes, each relying on its own specific trapdoor. We refer the reader to the survey [18] for details. We simply recall that the original proposal of Matsumoto and Imai [15] could be viewed as a multivariate variation on RSA and used as its internal MQ function a bijective monomial in some extension field. The resulting scheme was broken by Patarin in [16]. In order to repair the scheme, he later proposed the HFE cryptosystem [17], using a low degree polynomial as internal MQ function. In the same paper, he also proposed a wide range of variations on the HFE cryptosystem. We would like to remark that our result focuses on the basic scheme and that its extension to the variations is an interesting open problem.

As already mentioned, finding a solution of a generic system of MQ equations is NP-complete. For this reason, there were hopes that MQ schemes, HFE in particular, might very well be immune to subexponential attacks, square root attacks or even quantum computers. Indeed, these attacks are a common drawback of number theoretic cryptosystem and overcoming them is a worthy goal. To illustrate the possibilities, MQ schemes could achieve post-quantum computer security or yield extremely short signatures thanks to the lack of a square root attack. Such signatures could be, for a comparable security level, twice as short as a pairing based short signature of [3].

On the other hand, despite these hopes, several lines of attacks have been explored, such as the so-called relinearization techniques [12, 5], or the use of Gröbner basis algorithms [7]. The former was of a rather theoretical flavor, reducing the cryptanalysis of HFE to the resolution of an overdefined system of quadratic equations in the extension field with many excess equations, and describing a technique called relinearization to solve such overdefined systems in general. However the complexity of this attack remains unclear: despite the fact that relinearization was claimed to succeed in polynomial time, a close look at the claim shows that it only makes sense in a setting where the degree (called d in [17]) of the internal HFE polynomial is fixed. The latter approach followed a more experimental path since it used implementation of very generic Gröbner basis algorithms to solve the above quadratic system, thus breaking the first HFE Challenge 1 in 96 hours on a 833 MHz Alpha workstation with 4 Gbytes of memory. At a more abstract level, Faugère and Joux discovered an algebraic invariant that explains why the computation finishes much earlier than expected for a quadratic system of this size. Surprisingly, the authors did not try to derive a complexity bound for the problem of inverting HFE. This may be due to the lack of complexity estimates in the original paper [17] itself. This paper does not fully adhere to the current trend in cryptography, that defines a key generation algorithm with input a security parameter. In [17], the two main parameters of the schemes, the dimension n of the extension field and the degree d of the hidden polynomial are somehow unrelated. It is clear however that, in order to allow polynomial time decryption, any instantiation of HFE requires both n and d to be polynomial in the security parameter. Moreover, inverting HFE using exhaustive search has complexity 2^n. As a consequence, it is natural to use n itself as the security parameter. Thus d must be polynomial in n and we assume throughout the sequel that $d = O(n^\alpha)$, for some constant α.

This sheds some light on the hope that HFE might be immune to subexponential attacks. This hope stems from a remark in [17], which notes that the complexity of the so-called *affine multiple attack* is $O(n^{O(d)})$, thus exponential in the security parameter. Incidentally, this shows that the existence of polynomial time attacks for fixed d has been known right from the beginning. Similarly, the affine multiple attack is subexponential whenever d is small enough, say $d = O(\log n)$. In fact, in order to hope for full exponential security, we clearly need to choose $\alpha \geq 1$.

Another approach against MQ schemes uses the rank of the differential of the public key and has been proven successful to break the PMI scheme [11]. This technique also allows to build a quasipolynomial distinguisher for HFE [6], with complexity $O(\exp(c(\log n)^2))$, which happens to be the same as for our attack. As far as we know, this approach does not lead to a decryption attack against the HFE cryptosystem.

1.1 Our results

The main result of this paper is the following: there exists a heuristic quasipolynomial decryption attack against HFE. In fact, we do not actually propose a new attack but revisit the method described by Joux and Faugère in [7], which performs a Gröbner basis computation [4], with the efficient algorithms of Lazard and Faugère [14, 8, 9]. The efficiency of this approach was already shown by experiments from [7], and had been partly supported by mathematical arguments. Here, we give a more thorough theoretical analysis that allows to conclude that the attack is asymptotically efficient for any instantiation of HFE.

More accurately, our estimate yields complexity $O(n^{O(\log d)})$, that is $\exp(O(\log n)^2)$, for any instantiation of HFE where n is chosen as the security parameter, and where $d = O(n^\alpha)$, for some constant α. This greatly improves on the exponential estimate from [17]. This heuristic complexity estimate should also be compared with the complexity of factorization and integer discrete logarithm. In that case, we let n denote the size of the problem, i.e. $n = \log(N)$ to factor N or $n = \log(p)$ to compute discrete logarithm modulo p. The complexity is subexponential and its expression is $\exp(O(n^{1/3}(\log n)^{2/3}))$, which is clearly higher than what we obtain for HFE. In order to make this distinction clear, we say that the complexity of our attack is quasipolynomial rather than subexponential. Another widely used hard problem is the elliptic curve discrete logarithm problem, where the best attack has exponential complexity $O(\exp(n/2))$.

1.2 Organization of the paper

The paper is organized as follows: we first recall the definition of the HFE cryptosystem. Next, we survey known facts about Gröbner bases and their computation, focusing on the so-called degree of regularity of algebraic systems which is an extremely important parameter during the execution of Gröbner basis algorithms. Then comes the main contribution of the paper, where we bound the degree of regularity of the algebraic system arising from an attempt to directly invert HFE. This is done by showing that another system with a much smaller number of unknowns is in fact hidden into this algebraic system. Finally, we use this bound to show that the distinguishing and decryption attacks, obtained by applying a Gröbner basis computation to HFE systems, respectively have provable quasipolynomial and heuristic quasipolynomial complexities.

2 The HFE cryptosystem

Although the HFE cryptosystem was originally defined using any finite field as a base field, we restrict ourselves to the simpler case, where the base field is the two elements field.

2.1 Notations

Fields. We denote by $\mathbb{F}_2$ the finite field with two elements and by $\mathbb{F}_{2^n}$ the extension field with 2^n elements, which is isomorphic (as a vector space) to $(\mathbb{F}_2)^n$. A normal basis of $\mathbb{F}_{2^n}$ is defined by an element θ such that θ, θ^2, θ^4, $\ldots$, $\theta^{2^{n-1}}$ generate $\mathbb{F}_{2^n}$. It is well known that such a basis always exists. Note that the original description of HFE used a polynomial basis, however, since change of bases are linear and since arbitrary linear transforms are already used during the HFE construction, using a normal basis involves no loss of generality. Moreover, this approach greatly simplifies the exposition of our attack.

Monomials. Let $f(x) = x^d$ be a monomial over $\mathbb{F}_{2^n}$. The binary decomposition of the exponent reads $d = \sum_{0 \leq i < n} d_i 2^i$. Using the linearity of the Frobenius operator: $x \longrightarrow x^2$, it is easily seen that the Hamming weight of the d_i sequence is exactly the degree of the representation of f over $(\mathbb{F}_2)^n$, as a system of multivariate polynomial functions.

Polynomials. For any polynomial $P \in \mathbb{F}_{2^n}[X]$ we denote by $d^\circ P$ the degree of P, that is the maximal degree of its monomials. We let $w^\circ P$ be the maximal Hamming weight of the exponents of P's monomials, as defined in the previous paragraph, and call $w^\circ P$ the Hamming weight of P. It is well known that systems of MQ equations over $(\mathbb{F}_2)^n$ are in bijection with polynomials such that $w^\circ P = 2$. Also, affine functions over $(\mathbb{F}_2)^n$ are in bijection with polynomials such that $w^\circ P = 1$.

2.2 The HFE cryptosystem

The cryptosystem HFE is defined from a polynomial $f \in \mathbb{F}_{2^n}[X]$, with $w^\circ f = 2$ and $d^\circ f < d$, where n and d are (usually implicitly) defined from a security parameter. In the sequel, t will denote the smallest number such that $2^t > d$. The public key of HFE is obtained by composing f with two affine invertible functions, S, Y, thus yielding the polynomial $P = T \circ f \circ S$.

Encrypting with HFE is straightforward, it suffices to evaluate the public polynomial P on the input to be encrypted. Decryption is harder and uses the fact that it is easy to compute the inverses of S, T, and also to solve a polynomial equation of low degree $d < 2^t$ in time polynomial in d and n. Therefore, provided that $d = O(n^\alpha)$ and thus $t = O(\log n)$, a polynomial time decryption algorithm is available from the trapdoor.

3 Gröbner bases computations

Gröbner basis algorithms compute an algebraic basis of an ideal in a multivariate polynomial ring, given an ordering of the monomials. The output is such that any element f of the ideal can efficiently be written as an algebraic combination of the resulting basis $f_1, \cdots, f_m$ by repeating a sequence of simple reductions. Each reduction decreases the degree w.r.t. the ordering by suitably withdrawing from the current polynomial a multiple of some f_i by a monomial, until the zero polynomial is found. The original algorithm for computing Gröbner bases is due to Buchberger [4] and is based on maintaining a sequence of polynomials, and repeatedly using reduction, an operation that reduces the degree, and the so-called S-polynomial operation, an operation that increases the degree by computing an element of the ideal from a so-called critical pair of elements. In the early eighties, Lazard [13] realized that Gröbner basis computations could be achieved by applying Gaussian elimination to a specific matrix, called the Macaulay matrix, which is obtained by indexing the columns by all monomials with n variables of degree at most an integer r, and filling all rows with the coefficients of all multiples by a monomial of a family of polynomials generating the ideal, provided they remain of degree at most r. The main problem with this approach, is that the complete Macaulay matrix contains many "obvious" dependencies, which arise from generic properties and could be predicted in advance. Later Faugère [9] gave a simple criterion, that permit the construction of a reduced version of the Macaulay matrix that does not contain these obvious dependencies. This yielded a extremely efficient algorithm, called F_5, for the computation of Gröbner bases. In the sequel, we denote the variant of the Macaulay matrix containing all the multiples of degree r by the Macaulay-Faugère matrix of degree r.

The F_5 algorithm works by constructing Macaulay-Faugère matrices of increasing degree and by performing linear algebra on those. Its main goal is to find a linear combination of rows encoding a polynomial of degree smaller than r. By definition, the degree of regularity of a sequence $f_1, \cdots, f_m$ of polynomials is the minimal degree where such a linear combination exists. Each such linear combination encodes a new polynomial which needs to be added to the original sequence to get the Gröbner bases, except when the polynomial is zero in which case it might be necessary to remove some polynomial in the current ideal basis, thus simplifying it.

The degree of regularity D is a very important parameter of Gröbner basis computation using F_5, since it leads to a decomposition in two steps.

During the first step, up to the degree of regularity, the computation behaves nicely and its complexity can be easily predicted. What happens during the second step, when non-trivial combinations have appeared is much harder to predict in general. However, for random systems, the behavior is quite simple, an extremely large number of new polynomials appear in the Macaulay-Faugère matrices of degrees D or $D+1$, and after that the computation quickly terminates. Moreover, most real-life systems of equations have a similarly tame behavior and rarely need to construct Macaulay-Faugère matrices beyond the degree of regularity plus a small constant. On the other hand, it is possible to cook up wild systems with a very bad behavior.

For any system of polynomial equations in n unknowns with degree of regularity D, the first step of F_5 involves the construction of Macaulay-Faugère matrices up to degree D, thus of dimension at most n^D. Performing the linear algebra on these matrices costs at most n^{3D} operations. If we let $\mathcal{D}$ denote the largest degree of Macaulay-Faugère matrices occurring during the rest of the algorithm, the total cost is $n^{3\mathcal{D}}$. For well behaved systems of equations, $\mathcal{D}$ is not much larger than D and the overall complexity is $n^{O(D)}$.

3.1 Known bounds on degrees of regularity

Previous work shows that for a quadratic system of equations in n variables, the degree of regularity cannot become too large. Moreover, for random systems of equations, the known bound on the degree of regularity is reached. The general analysis that we need was done by Bardet, Faugère and Salvy [2] and is neatly described in Bardet's thesis [1, chap. 4].

The result given there is that for a system of τn quadratic equations in n unknowns, the degree of regularity is at most:

$$D_\tau(n) = \left(\tau - \frac{1}{2} - \sqrt{\tau(\tau-1)}\right) n + \frac{-a_1}{2(\tau(\tau-1))^{1/6}} n^{1/3} \tag{1}$$

$$- \left(2 - \frac{2\tau - 1}{4\sqrt{\tau(\tau-1)}}\right) + O(n^{-1/3}), \tag{2}$$

where $a_1 \approx -2.33811$ is the first real zero of the Airy function.

Moreover, for random systems the probability of having a smaller degree of regularity is negligible. Furthermore, experiments show that such random systems are well behaved in the sense that the F_5 computation do not involve Macaulay-Faugère matrices of much higher degree. To formalize this observation, we propose the following conjecture:

Conjecture 1. For all $\tau > 1$, there exists a constant K such that for a large enough random system S of $\lceil \tau n \rceil$ quadratic equations in n unknowns, a Gröbner basis for the ideal generated by S can be computed in time n^{Kn}, with overwhelming probability.

4 Systems of equations arising from HFE instances

In this section, our goal is to study the complexity of a direct Gröbner basis approach to the resolution of HFE systems. This direct approach consists in writing down that each public polynomial, belonging to the encryption key, when evaluated on the (unknown) plaintext yields the corresponding ciphertext bit. This approach was first described in [7].

4.1 Outline of the strategy

Our strategy is to bound the degree of regularity D of the system of polynomials stemming from directly attempting to invert HFE from the description of its public key through a Gröbner basis algorithm. From such a bound, which is smaller than the bound for a random system in the same number of unknowns, we then derive our main results. First, when the first simplification in a Macaulay-Faugère matrix is encountered, we deduce the effective degree of regularity of the polynomial system under consideration. If this degree is small enough, our distinguisher knows that the system has no chance to be random and asserts that it is an HFE instance. No heuristic is required for this attack, whose runtime is bounded by n^{3D}. Second, for the decryption attack, we assume that HFE based systems of equations behave nicely and that $\mathcal{D}$ is not much larger than D. Under this heuristic assumption, which is supported by the experiments described in [10, 7, 1], where the computations never constructed Macaulay-Faugère matrices beyond the degree of regularity plus 1, we claim that the complexity of the decryption attack remains $n^{O(D)}$.

The key idea that allows us to bound D is to apply a sequence of transformations involving the unknown secret trapdoor, in order to show that D does not exceed the degree of regularity of a much smaller system. More precisely, this other system involves $l+t+l-2 = (2\lambda+1)\cdot t-2$ equations in $(\lambda+1)\cdot t$ unknowns, where 2^t is, as defined in section 2.2 a bound for the degree of the internal HFE polynomial and λ is an appropriate constant. Since this system has a much smaller number of unknowns, we obtain a much better bound on D than for generic systems in n unknowns.

4.2 Reducing the number of unknowns

Let θ, θ^2, θ^4, $\ldots$, $\theta^{2^{n-1}}$ be our normal basis for the finite field $\mathbb{F}_{2^n}$. Let $f(X)$ be the secret polynomial of an HFE instance over $\mathbb{F}_{2^n}$, of degree $d < 2^t$. The corresponding public key polynomials in n unknowns $x_0, \ldots, x_{n-1}$ over $\mathbb{F}_2$ are P_1, P_2, $\ldots$, P_n. They are obtained by writing X over the normal basis as:

$$X = \sum_{0 \le i < n} \theta^{2^i} x_i,$$

by taking the coordinates of $f(X)$, viewed as polynomials in x_0, $\ldots$, x_{n-1}, in the normal basis and, finally, by applying two invertible linear transforms, as explained in section 2.2.

The resulting polynomials P_1, P_2, $\ldots$, P_n are quadratic. Thus, cryptanalyzing a message encrypted with the HFE cryptosystem requires solving a quadratic system of equations over $\mathbb{F}_2$ defined by fixing the target values of the f_i polynomials. In order to decrypt an HFE instance using a generic Gröbner basis approach, this is the system of equations we need to consider. Thus, following our general strategy, we want to bound its degree of regularity.

First of all, since the degree of regularity only depends on the high degree homogeneous parts of each equation in the system, the target value has no influence on this parameter. Moreover, assuming that the quadratic parts of the P_i, together with the quadratic parts of the field equations $x_i^2 - x_i = 0$, are linearly independent, the degree of regularity remains the same if we remove the secret linear transformations before and after f. At this point, we are left with computing (or more precisely bounding) the degree of regularity of the system of secret internal equations directly given by the coordinates of f over the normal basis, together with the field equations $x_i^2 - x_i = 0$. Let f_0, f_1, $\ldots$, f_{n-1} denote these secret polynomials, which are related to f by means of the equation:

$$f\Big(\sum_{0 \le i < n} \theta^{2^i} x_i\Big) = \sum_{0 \le i < n} \theta^{2^i} f_i.$$

Another system with higher degree of regularity than the internal system. In order to bound the degree of regularity of the ideal generated by $(f_0, \cdots, f_{n-1}, x_0^2 - x_0, \cdots, x_{n-1}^2 - x_{n-1})$, we first remark that this degree is left unchanged when solving this system over $\mathbb{F}_{2^n}$ instead of $\mathbb{F}_2$. Moreover, thanks to the field equations, the solutions are the same in both cases. Over $\mathbb{F}_{2^n}$, we can transform the system by writing:

$$F_j(x_0, \cdots, x_{n-1}) = \sum_{0 \le i < n} \theta^{2^{i+j}} f_i.$$

In fact, F_0 is just a representation of f in terms of $x_0, \ldots, x_{n-1}$, F_1 is a representation of f^2, F_2 a representation of f^4 and so on ...

Clearly, replacing the f_i by the F_j is an invertible linear transform, which for the same reasons as before, does not affect the degree of regularity. The next step is to make a linear change of variables, replacing the x_i by y_j, where:

$$y_j = \sum_{0 \le i < n} \theta^{2^{i+j}} x_i.$$

This change corresponds to setting $y_0 = X$, $y_1 = X^2$, $\ldots$, $y_{n-1} = X^{2^{n-1}}$. It cleanly expresses each F_i as a quadratic equation in terms of X and its Frobenius images. However, the field equations are not yet in a nice form. Luckily, a final linear transform turns them into:

$$\begin{aligned} y_1 &= y_0^2, \\ y_2 &= y_1^2, \\ &\vdots \\ y_{n-1} &= y_{n-2}^2, \\ y_0 &= y_{n-1}^2. \end{aligned}$$

Finally, in order to bound the degree of regularity, it is enough to remark that due to the degree bound on f, F_0 is a function of $y_0, \ldots, y_{t-1}$ and that $y_t, \ldots, y_{n-1}$ are not used. Likewise, F_1 is a function of $y_1, \ldots, y_t$ and F_j a function of $y_j, \ldots, y_{t+j-1}$ (when j is small enough). Thanks to this observation, we can focus on a subset of the equations and variables. Assume that we restrict ourselves to $F_0, F_1, \ldots, F_{l-1}$ then we need only use the variables y_0 to y_{t+l-1}. Moreover, among these variables, we keep $t+l-2$ field equations of the form $y_{j+1} = y_j^2$. Of course, restricting ourselves to such a subset can only increase the degree of regularity, since any non trivial relation among the equations of the smaller system clearly holds in the larger one. Setting $l = \lambda t$, for an adequately chosen constant λ, we now obtain a system of $l + t + l - 2 = (2\lambda + 1) \cdot t - 2$ equations in $(\lambda + 1) \cdot t$ unknowns.

Note. In the case where $d = 2^t$, we can slightly improve the above description to reduce the complexity of the attack. This is due to the fact that the variable y_t in F_0 only appears as a linear term. Thus, the additional variable y_{t+l} only appears linearly. Since the degree of regularity only depends on the high order (quadratic) terms, we can safely ignore this extra variable. This confirms the practical observation made in [7], in connection with the degree of the polynomial complexity with fixed degree d. They observed that this degree slowly increased with d and that the increase steps occurred immediately after increasing d beyond a power of 2.

4.3 Bound on the degree of regularity of the internal system

To upper bound the degree of regularity arising with HFE systems, it is now sufficient to apply the generic bound for random systems on the internal system with a reduced number of variables, expressed as a function of $t = \lceil \log_2 d \rceil$ as in section 4.2.

We apply the bound from section 3.1 with $\tau = (2\lambda+1)/(\lambda+1)$, fixing $\lambda = 1$. This is not the optimal choice and it would be better to let λ grow with n in order to make τ close to its limit 2, thus finding a tighter bound. However, the simple choice $\lambda = 1$ is sufficient to fulfill our purpose. With this choice, we find that ignoring low order terms the degree of regularity of $3t$ quadratic equations in $2t$ unknowns is:

$$D = 2(1 - \sqrt{3/4})t + O(t^{1/3}).$$

This can be summarized by the following theorem:

Theorem 1. *For basic HFE instances, defined by a secret polynomial of degree d over $\mathbb{F}_{2^n}$, the maximum possible degree of regularity $D(n,d)$ is asymptotically upper bounded by*

$$(2+\epsilon)(1 - \sqrt{3/4}) \min(n, \log_2 d),$$

for all ϵ.

4.4 Complexity of the attacks

The above study has shown that the degree of regularity of the HFE system is an integer D upper bounded by $2(1 - \sqrt{3/4})t + O(t^{1/3})$. As noted in section 3, the complexity of computing a Gröbner basis with F_5 is bounded by the cost of linear algebra on a matrix whose columns are indexed by the monomials of degree $\mathcal{D}$ in n unknowns, where $\mathcal{D}$ is the degree of the largest Macaulay-Faugère matrix that is used. Moreover, during a first phase the algorithm only builds Macaulay-Faugère matrices of degree up to D. Moreover, the end of the first phase is easily detected. As a consequence, we easily measure the effective value of D. Since we do not know the HFE secret key, we need to perform our Gröbner basis computation in the public world, using a system with n unknowns. Thus the respective runtimes of the first phase and of the full algorithm are bounded by $O(n^{3D})$ and $O(n^{3\mathcal{D}})$, assuming that all linear algebra is done using ordinary Gaussian elimination.

An alternative way of viewing this result is to make the heuristic assumption that the reduced system described in section 4.2 behaves as a random system in t unknowns. Thus using conjecture 1, a Gröbner basis for this hidden system can be found in time t^{Kt}. Further assuming that this can be achieved with the F_5 algorithm, the Gröbner computation in the public world requires a running time n^{Kt}.

Distinguishing attack At the end of the first phase, we end up with the effective value of D. We know that for HFE instances, this value is quite small and that for random systems, it is much larger (with overwhelming probability). This simple fact yields a simple distinguisher which at the end of the first phase can tell whether the original system is an HFE instance or not. Of course, when working on a random system, the distinguisher should use an early abort strategy and stop as soon as the current Macaulay-Faugère matrix has degree larger than the expected D. Thanks to the early abort strategy, replacing D by its bound, we find the complexity of the distinguisher is $2^{O(\log(n)^2)}$ even when the input is a random system. This distinguisher was first mentioned by Faugère when describing his HFE challenge experiment. It offers an alternative to the distinguisher described in [6].

Note that for a random system of quadratic equations, we need a different variant of the theorem of Bardet than described in section 3.1. This variant, which holds for quadratic systems of n equations in n unknowns over $\mathbb{F}_2$, is also given in [1] gives a formula for the degree of regularity which is similar to equation 2 for $\tau = 2$, with slightly different constants. It implies that the degree of regularity of a random system is $O(n)$ and thus much higher than the $O(\log n)$ value we obtained for D in the case of an HFE instance.

Decryption attack In the case of the decryption attack, we let the Gröbner basis algorithm terminate and from the result, we find the corresponding plaintext as in [7]. Here the runtime is $n^{3\mathcal{D}}$. Using our heuristic assumption about the good behavior of the system of equations, $n^{3\mathcal{D}}$ is $n^{O(D)}$.

As a consequence, the overall heuristic runtime is $2^{O(\log(n)^2)}$ as announced previously.

5 Conclusion

In this paper, we analyzed the behavior of the Gröbner basis based attack on HFE systems as proposed in [7]. We showed that this attack takes quasipolynomial time for any practical instantiation of the basic HFE cryptosystem. The runtime analysis of the distinguisher part of the attack gives a provable complexity, while the decryption part of the attack only leads to a heuristic complexity. Comparing the result with the best existing subexponential algorithms for factoring and computing discrete logarithms, we find that for comparable key and/or ciphertext sizes, breaking the basic HFE scheme is asymptotically much easier than breaking RSA or discrete logarithm based systems.

References

1. M. Bardet. *Étude des systèmes algébriques surdéterminés. Applications aux codes correcteurs et à la cryptographie.* PhD thesis, Université Paris 6, Dec. 2004. `http://www-calfor.lip6.fr/~bardet/`.
2. M. Bardet, J.-C. Faugère, and B. Salvy. On the complexity of Gröbner basis computation of semi-regular overdetermined algebraic equations. In *Proc. ICPSS International Conference on Polynomial System Solving*, 2004.
3. D. Boneh, B. Lynn, and H. Shacham. Short signatures from the Weil pairing. In C. Boyd, editor, *Proceedings of ASIACRYPT'2001*, volume 2248 of *Lecture Notes in Comput. Sci.*, pages 514–532. Springer, 2001.
4. B. Buchberger. Gröbner bases : an algorithmic method in polynomial ideal theory. In N.-K. Bose, editor, *Multidimensional systems theory*, number 16 in Mathematics and its Applications, chapter 6, pages 184–232. D. Reidel Pub. Co., 1985. Based on his PhD thesis: *Ein Algorithmus zum Auffinden der Basiselemente des Restklassenringes nach einem nulldimensionalen Polynomideal*, U. Innsbruck, Austria, 1965.
5. N. Courtois. The security of Hidden Field Equations (HFE). In *CT-RSA'01*, volume 2020 of *Lecture Notes in Comput. Sci.*, pages 266–281. Springer-Verlag, 2001.
6. V. Dubois, L. Granboulan, and J. Stern. An efficient provable distinguisher for HFE. In *Proceedings of ICALP*, Lecture Notes in Comput. Sci., 2006. To appear.
7. J.-C. Faugère and A. Joux. Algebraic cryptanalysis of Hidden Field Equation (HFE) cryptosystems using Gröbner Bases. In *Crypto'03*, volume 2729 of *Lecture Notes in Comput. Sci.*, pages 44–60. Springer-Verlag, 2003.
8. J.-C. Faugère. A new efficient algorithm for computing Gröbner bases (F4). *J. Pure Appl. Algebra*, 139(1-3):61–88, 1999. Effective methods in algebraic geometry (Saint-Malo, 1998).
9. J.-C. Faugère. A new efficient algorithm for computing Gröbner bases without reduction to zero (F5). In T. Mora, editor, *ISSAC 2002*, pages 75–83, 2002.

10. J.-C. Faugère. Algebraic cryptanalysis of HFE using Gröbner bases. Technical Report 4738, INRIA, Feb. 2003. `ftp://ftp.inria.fr/INRIA/tech-reports/dienst/RR-4738.pdf`.
11. P.-A. Fouque, L. Granboulan, and J. Stern. Differential cryptanalysis for Multivariate Schemes. In *Eurocrypt'05*, volume 3386 of *Lecture Notes in Comput. Sci.*, pages 341–353, 2005.
12. A. Kipnis and A. Shamir. Cryptanalysis of the HFE Public Key Cryptosystem. In *Crypto'99*, volume 1666 of *Lecture Notes in Comput. Sci.*, pages 19–30. Springer-Verlag, 1999.
13. D. Lazard. Gröbner bases, gaussian elimination and resolution of systems of algebraic equations. In *Computer algebra (London, 1983)*, volume 162 of *Lecture Notes in Comput. Sci.*, pages 146–156, 1983.
14. D. Lazard. Solving systems of algebraic equations. *ACM SIGSAM Bulletin*, 35(3):11–37, Sept. 2001.
15. T. Matsumoto and H. Imai. Public Quadratic Polynomial-tuples for efficient signature-verification and message encryption. In *Eurocrypt'88*, volume 330 of *Lecture Notes in Comput. Sci.*, pages 419–453. Springer-Verlag, 1988.
16. J. Patarin. Cryptanalysis of the Matsumoto and Imai Public Key Scheme of Eurocrypt'88. In *Crypto'95*, volume 963 of *Lecture Notes in Comput. Sci.*, pages 248–261. Springer-Verlag, 1995.
17. J. Patarin. Hidden Field Equations (HFE) and Isomorphisms of Polynomials (IP): two families of asymetric algorithms. In *Eurocrypt'96*, volume 1070 of *Lecture Notes in Comput. Sci.*, pages 33–46. Springer-Verlag, 1996.
18. C. Wolf and B. Preneel. Taxonomy of Public Key Schemes based on the problem of Multivariate Quadratic equations. Cryptology ePrint Archive, Report 2005/077, 2005. `http://eprint.iacr.org/`.