## **CryptoChallenge 11**

Jean-Charles Faugère SALSA/INRIA Rocquencourt LIP6/CNRS Université Paris 6 (UPMC) case 168, 4 pl. Jussieu, F-75252 Paris Cedex 05 E-mail: Jean-Charles.Faugere@lip6.fr June 10, 2005

#### Abstract

In this short report we present an efficient attack of the  $C^*$  cryptosystem based on fast algorithms for computing Grbner basis. The attack consists in computing a Grbner basis of the public key. The efficiency of this attack depends strongly on the choice of the algorithm for computing the Grbner basis: it was was possible to break the crypto-challenge 11 in only 3 hours and 11 minutes of CPU time (PC Pentium 2.8 Ghz Xeon) by using the algorithm  $F_5$  implemented in C. We recommend to increase the value of  $n \ge 26$ .

## **1** Description

## **1.1** The $C^*$ crypto system

Let  $q = 2^m$ , D = 4, K is the finite field GF(q) and  $L = GF(2^{nm})$  an extension field of K.

$$\pi(x) = x^d, \ d = \sum_{k < D} q^{r_k} \quad r_0 < r_1 < \dots < r_{D-1} < n$$

**Proposition 1**  $\pi(x)$  is bijective on K if and only if  $gcd(d, q^n - 1) = 1$ 

The private key consists of two bijective K-affine mappings S and T on L, each represented by a non-singular  $n \times n$  matrix over K and a vector in  $K^n$ . The encryption is defined as

$$E(x) = S(\pi(T(x))), \quad x \in L$$

Knowing S and T allows to decrypt, since one-to-one affine mappings and power functions can easily be inverted. The public key is the multi-variate representation of E(x) on  $K^n$ , leading to a collection of n polynomials (of total degree 4):

$$(Pub) \quad E_i(x_0, \dots, x_n) \ i = 0, \dots, (n-1)$$

## 1.2 Challenge 11

The challenge 11 is a particular instance of the  $C^*$  cryptosystem: m = 5 and n = 16 and one has to solve the following algebraic system

$$(S)$$
  $E_i(x_0, \dots, x_n) = Q_i, \ i = 0, \dots, (n-1)$ 

where  $Q_i$  is explicitly given:  $(Q_0, \ldots, Q_{n-1}) = (w^2 + 1, w^2 + 1, w^3 + w, w^4 + w^3 + w^2, w^2 + 1, w^4 + w^3 + w + 1, w^3 + w, w^3 + w^2 + w, 1, w^4, w^3 + w^2 + w + 1, w^2 + w + 1, w^4 + w^3 + 1, w, w^4 + w^3 + w^2 + w + 1) \in K^n$ 

and  $E_i$  is a polynomial of degree 4 in  $x_0, \ldots, x_{n-1}$  and coefficients in K. To "break" the challenge we have to compute

$$\mathcal{V}_K = \{ (x_i) \in K^n \, | \, E_i(x) = Q_i \}$$

**Proposition 2**  $\mathcal{V}_K$  is of dimension 0 and degree 1.

**Proof** From the proposition 1.  $\Box$ 

## 2 Gröbner bases attack

### 2.1 Solutions in the ground field

If  $\bar{K}$  is the algebraic closure of K then a Gröbner basis computation of S gives a description of

$$\mathcal{V}_{\bar{K}} = \{(x_i) \in \bar{K}^n \,|\, E_i(x) = Q_i\}$$

### **Experimental Fact 1**

m	n	$degree(\mathcal{V}_{\bar{K}})$	$degree(\mathcal{V}_K)$
5	8	32	1
5	9	76	1
5	10	88	1
5	11	4	1
5	12	112	1
5	13	628	1
5	14	568	1
5	15	5324	1
5	16	6208	1
6	8	32	1
6	10	88	1
6	14	568	1

We deduce from the previous experiments that computing directly the Gröbner basis of S contains parasite solutions.

## 2.2 First attack

To force the solutions to be in K we can add the "field equations"  $x_i^q = x_i$  but since q = 32 this is useless in this case. The other solutions is to give a value to one (or several) variable:

#### **Algorithm 1st attack**

for *i* from 0 to m - 2 do Substitute  $x_n = w^i$  in SCompute  $G_i$  a Gröbner basis of this system if  $G_i \neq \{1\}$ Add the field equations  $x_j^q = x_j$  to  $G_i$ Compute  $G'_i$  a Gröbner basis of this system This is the solution of the algebraic system.

Of course all the computations can be computed in parallel of different computers. In our case, we used the  $F_4$  algorithm we found a solution for n = 20:

 $\begin{array}{l} x_1 \ = \ w^{27}, x_0 \ = \ w^{16}, x_2 \ = \ w^{18}, x_3 \ = \ w^{17}, x_4 \ = \ w^{25}, x_8 \ = \ w^{16}, x_9 \ = \ w^{22}, x_{10} \ = \ w^{28}, x_{11} \ = \ w^{12}, x_{15} \ = \ w^{20}, x_{14} \ = \ w^{18}, x_{13} \ = \ w^{16}, x_{12} \ = \ w^{11}, x_5 \ = \ w^4, x_7 \ = \ w^{29}, x_6 \ = \ w^{29} \end{array}$ 

It takes about 3 hours of CPU (Intel Pentium Xeon 2.8 Ghz) to compute the Gröbner basis  $G_i$ . Thus the total sequential time is  $32 \times 3 = 96$  hours of computation and 3 hours of parallel CPU time.

### 2.3 New attack

We use a different strategy: compute the Gröbner basis of the whole system S using the  $F_5$  algorithm; in a second step (the fastest part of the computation) we select the solutions in  $K^n$ . The computation can be carried out on a single PC (2.8 Ghz) with 2Go bytes of memory and it takes 11473.78 sec to compute the Gröbner basis; the number of solution is 6208 in  $\bar{K}^n$  and the size of the Gröbner basis is 266 Mbytes.

## **3** Complexity of the attack

It takes about 3 hours of CPU (Intel Pentium Xeon 2.8 Ghz) to compute the Gröbner basis. For the challenge 1/HFE the total CPU is about 48 hours (Alpha DS25 1 Ghz).

However to compare more precisely this is interesting to compare two parameters:

- The number of arithmetical operations.
- The maximal degree occurring in the computation of the Gröbner bases.

### **3.1** Number of operations

The total number of operations (XOR with 64 bits words) to break HFE challenge 1 is about  $2^{44.46}$ .

The total number of operations (multiplication if GF(32)) to break  $C^*$  challenge 11 is about  $2^{41.2}$  in less that 3 hours 11 mins.

## 3.2 Maximal degree

The maximal degree occurring in the computation of a Grbner basis is a very important since

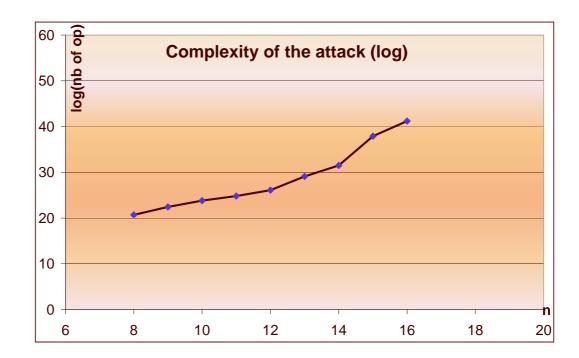
- It is an estimate of the non randomness of an algebraic system. The regularity D of a generic system is given by the Macaulay bound and in our case we should have: 1 + ∑<sub>i=1</sub><sup>n</sup>(d<sub>i</sub> − 1) = 3n + 1
- It gives a (rough) estimate of the complexity of the Grbner basis computation:  $O\left(\binom{n+D}{n}^{\omega}\right)$  where  $N^{\omega}$  is the cost of the multiplication of two  $N \times N$  matrices.

d	16	17	33	96	128	129	257	384	512	513
Max degree	3	4	4	4	4	5	5	5	5	6
Maximal degree for HFE, $d = 2^i + 2^j \operatorname{GF}(2)$										

From the previous table we see that when d (degree of the univariate polynomial) is fixed the corresponding HFE can be solved in polynomial time (see also the paper Faugère/Joux Crypto 2003).

n	8	9	10	11	12	13	14	15	16
Max degree	6	7	7	6	6	6	8	8	8
Nb of *	$2^{20.7}$	$2^{22.4}$	$2^{23.8}$	$2^{24.8}$	$2^{26.1}$	$2^{29.1}$	$2^{31.5}$	$2^{37.9}$	$2^{41.2}$
Maximal degree for $C^*$ , $d = 1 + q + q^5 + q^7$ , $q = 2^5 = 32$									

egree for  $C^*$ ,  $d = 1 + q + q^5 + q^4$ , qNew attack using  $F_5$ 



n	8	10	12	13	14	15	16	18
Max degree	6	6	6	6	7	7	7	7 or 8
Nb of * (parallel)	$2^{20.7}$	$2^{23.8}$	$2^{26.1}$	$2^{29.2}$	$2^{30.0}$	$2^{37.8}$		
Nb of * (sequent)	$2^{25.7}$	$2^{28.8}$	$2^{31.1}$	$2^{34.2}$	$2^{35.0}$	$2^{42.8}$		
Maximal degree for $C^*$ , $d = 1 + q + q^5 + q^7$ , $q = 2^5 = 32$ ,								
Specialise one variable.								

First attack using  $F_4$ 

**Remark 1** Note that for n = 18,  $gcd(d, q^n - 1) = 19$ . The test has been done by specialising 2 and 3 variables.

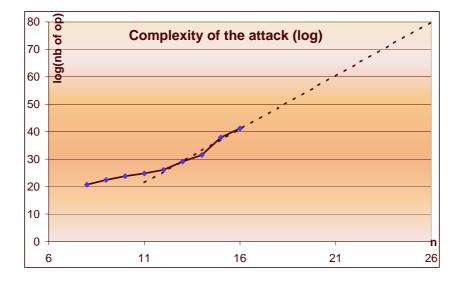
We see from the previous table that  $C^*$  perhaps can be solved in polynomial time when d is fixed since the maximal degree occurring in the computation does not depends on n. Of course a mathematical proof is needed.

**Experimental Fact 2** The complexity does not depend on m > 2 the size of the finite field  $GF(2^m)^1$ 

**Proof** This is the the result of experiments for m = 3, 4, 5, 6, 7. For m = 2 it is necessary to add the field equations  $x_i^4 - x_i$  and the Gröbner computations behave differently.  $\Box$ 

# 4 Conclusion

If we extrapolate the previous results we found:



The CryptoChallenge 11 can	be broken easily $2^{41.1}$ (about 3 hours )
To improve the	security of $C^*$ it seems that
	Increase the size of $K = GF(2^m)$ is useless
	Increase the value $n \ge 26$ (?) can lead to a
	much more difficult algebraic system.

Despite a substantial amount of real computer simulations, *major uncertainties* remain.

<sup>&</sup>lt;sup>1</sup>In fact the complexity depends on m but slightly.

## MYSTERY TWISTER 2005 www.mystery-twister.com

CRYPTOCHALLENGE 11

April 2005

## Multi-variate Public Key Schemes: Biquadratic C\*

Public key cryptosystems based on multi-variate polynomials are studied since the eighties. One of them, called C<sup>\*</sup>, survived for some years, while most were broken in a very short time after being proposed. C<sup>\*</sup> was introduced in 1988 by Imai and Matsumoto, and it was broken in 1994. But this is not the end of the story. The C<sup>\*</sup> design, suitably modified, still provides interesting objects for cryptographic investigations. In the sequel we shall explain why.

#### **Definition of** $C^*$

As basic public parameters of C<sup>\*</sup>, a finite field  $K = GF(2^m)$ , an extension field  $L = GF(2^{nm})$  of K, and a bijective power function

$$\pi(x) = x^d, \quad d = 1 + q^r, \quad (q := 2^m)$$

are chosen. L is identified with  $K^n$ . Recall that  $x^d$  is bijective on L if and only if  $gcd(d, q^n - 1) = 1$ .

The private key consists of two bijective K-affine mappings S and T on L, each represented by a non-singular  $n \times n$  matrix over K and a vector in  $K^n$ . The encryption is defined as

$$E(x) = S(\pi(T(x))), \quad x \in L.$$

Knowing S and T allows to decrypt, since one-to-one affine mappings and power functions can easily be inverted. The public key is the multi-variate representation of E(x) on  $K^n$ , leading to a collection of n quadratic polynomials

$$E_i(x_0, ..., x_{n-1}), \quad i < n,$$

over K with variables  $x_k$  (k < n). The idea of C<sup>\*</sup> is that S and T are hidden in the coefficients of these multi-variate polynomials.

In fact it remains an open challenge to recover the private key (S, T) from the public key, i.e. the  $(E_i)_{i < n}$ .

### Breaking $C^*$

Independently Dobbertin (1994, unpublished) [1] and Patarin (1995) [4] found the same "linearization" attack. It does not find the private key but a kind of substitute, such that each encryption comes down to solving a linear equational system over K with n unknowns:

The inversion of E(x) can be reduced to a linear problem. To confirm this we can assume w.l.o.g. for a moment that S and T are the identity mappings. Suppose  $a = x^{1+q^r}$ , then

$$ax^{q^{2r}} + a^{q^r}x = 0, (1)$$

is a K-linear equation over L for each given  $a \in L$ . It is a routine matter to show that the K-rank of (1) is  $n - \gcd(r, n)$  for non-zero a.

Without knowing S and T we can derive the general pattern of the n linear equations over K underlying the corresponding affine modification of (1) by computing the vector space

$$\mathcal{K} = \{\lambda(x) \in L[x] : \lambda_i(x), \, i < n, \, \text{and} \, \sum_{i < n} \lambda_i(x) E_i(x) \text{ are affine}\}, \qquad (2)$$

where the  $\lambda_i$  and  $E_i$  are considered as multi-variate polynomials. The computation of  $\mathcal{K}$  is a linear problem. Choosing a base of  $\mathcal{K}$  we can decrypt. Suppose that Q = E(P) is given and P has to be found. Then for each  $\lambda \in \mathcal{K}$  we get an affine equation

$$\sum_{i < n} \lambda_i(x_0, ..., x_{n-1}) E_i(x_0, ..., x_{n-1}) = \sum_{i < n} Q_i \lambda_i(x_0, ..., x_{n-1}),$$

which is satisfied by x = P. It turns out that usually decryption with  $\mathcal{K}$  is faster than decryption with the private key!

#### Stronger modifications of C\*

As a modification of C<sup>\*</sup>, which avoids the described devastating weakness of the original system, Patarin proposed HFE (Hidden Fields Equations) cryptosystems [5], where the power function  $\pi$  is replaced by a low degree polynomial

$$p(x) = \sum_{k} a_k x^{q^{r_k} + q^{s_k}}$$

over L. The form of p implies that the representation of p on  $K^n$  leads again to *quadratic* multi-variate polynomials. However, the replacement of  $\pi$  by an in general non-bijective p causes a lot of difficulties. It makes decryption (or signing if HFE is used for digital signatures) rather complicated. To determine all P with E(P) = Q, one has to compute all zeros of p(x) + c with  $c = S^{-1}(Q)$ . Moreover, to identify the proper P it is necessary to mark E(P) with a hash value of P. This effort means paying a high prize for trying to make C<sup>\*</sup> secure. – We note that there is a zoo of HFE variants [5]. We consider here only the basic version.

In 2002 HFE was broken for 80 bit block size by Faugère [3] in the most difficult case K = GF(2) (HFE Challenge 1, see [5]). He used powerful elimination techniques for multi-variate equational systems.

Another modification of C<sup>\*</sup> was studied in [1] (unpublished). Here the exponent  $d = 1 + q^r$  is replaced by

$$d = \sum_{k < D} q^{r_k}, \quad r_0 < r_1 < \dots < r_{D-1} < n,$$
(3)

for some fixed D > 2, were  $r_0 = 0$  can be assumed without loss of generality. Note that the degree of the associated  $E_i$  is D. Hence for practical applications the restriction D = 3 or D = 4 makes sense, since otherwise the size of the  $E_i$ becomes too large.

#### The challenge

CRYPTOCHALLENGE 11. A biquadratic variant of the  $C^*$  scheme with m = 4, n = 25 and

$$d = 1 + q + q^3 + q^{12} \quad (q = 16)$$

is considered. Some private key (S,T) is chosen. Thus an asymmetric cryptosystem is defined, which operates on 100 bit blocks.

The encryption Q of an 100 bit block P is given. Decrypt Q based on the knowledge of the public key, i.e. the multi-variate polynomials  $E_i(x_0, ..., x_{24})$ , i < 25, over  $K = GF(2^4)$ . This means that one has to compute P by solving the system of the biquadratic equations  $E_i(x_0, ..., x_{24}) = Q_i$ , i < 25, over K. (See below for the further technical specification of this challenge.)

#### HFE vs. biquadratic $C^*$

We emphasize that we do not propose to apply biquadratic  $C^*$  with the parameters in CC11 in practice. The parameters are chosen at the edge, implying the risk of being broken. The intention of this challenge is

- to stimulate research on solving systems of multi-variate equations and
- to compare the cryptographic strength of a biquadratic C<sup>\*</sup> and a HFE system of the same block size. (In a broader field of investigations also cubic C<sup>\*</sup> variants should be studied.)

The most obvious advantage of biquadratic  $C^*$  is that

• decryption (signing) remains as simple as for the original C<sup>\*</sup>.

For  $C^*$  system of degree D we have the formula

block size:	nm bit,
public key length:	$\binom{n+D}{D}nm$ bit,
private key length:	2mn(n+1) bit.

Up to the specification of a low degree polynomial, we have the same parameters for (quadratic)  $C^*$  and HFE.

For the system in CRYPTOCHALLENGE 11 we have

block size:	100 bit,
public key length:	$290~{\rm kb},$
private key length:	5,200 bit.

For practical applications such relatively long public keys are usually not a serious problem. However, the private key should be short, since it has to be stored and used in a secure environment, often with limited resources (smart cards).

For a counterpart to the biquadratic C\* version in CC11, the HFE system with m = 1 and n = 100, we have

block size:	100 bit,
public key length:	63  kb,
private key length:	20,200 bit.

It would be interesting to compare the cryptographic strength of the latter HFE system with the system in CC11.

### Weak exponents

In this section we cite results from [1]. (Proofs can be found in [2].) We shall refer to the following generalization of (2) for  $d = \sum_{k < D} q^{r_k}$ ,  $r_0 < r_1 < ... < r_{D-1} < n$ , and arbitrary degree r:

$$\mathcal{K}^{(r)}(d) = \{\lambda(x) \in L[x] : \deg(\lambda_i(x)) \le r \ (i < n), \ \deg\left(\sum_{i < n} \lambda_i(x) E_i(x)\right) \le r\},\$$

**Theorem 1** For any exponent d we have

$$\dim_K \mathcal{K}^{(D-1)}(d) \ge \binom{D}{2} n.$$

In general we have  $\mathcal{K}^{(r)}(d) = 0$  for r < D - 1. Otherwise we call d weak, since then we get a C<sup>\*</sup> variant which has certain weaknesses against elimination attacks using Gröbner base techniques. But for all exponents d Theorem 1 implies that starting with the given equations of degree D, increasing the degree of derived equations up to 2D - 1 we have a "degeneration effect" and can get equations of degree D - 1, independent of K. (The special structure of C<sup>\*</sup> for arbitrary degree D allows to compute the latter degree D - 1 equations a priori simply by computing  $\mathcal{K}^{(D-1)}$ . This works precisely as described before in the case D = 2 for the breaking of classical C<sup>\*</sup>.)

**Theorem 2** An exponent d is weak if there are  $k, \ell, u, v < D$  such that  $k \neq \ell$ ,  $u \neq v, (k, \ell) \neq (u, v)$  and

$$r_k - r_\ell = r_u - r_v \bmod n. \tag{4}$$

For D = 4 it is obvious that this implies  $n \ge 13$ , since otherwise all exponents are weak. (In fact there are precisely 12 pairs  $(k, \ell)$  with  $k \ne \ell$  in  $\{0, 1, 2, 3\}^2$ , and therefore the mapping  $(k, \ell) \mapsto r_k - r_\ell \mod n$  into  $\{1, 2, ..., n - 1\}$  must have collisions for  $n \le 12$ .) We anticipate that the converse of Theorem 2 is also valid.

The weakness becomes extremal if d is a geometric series:

**Theorem 3** For exponents of the form  $d = 1 + q^r + q^{2r} + ... + q^{(D-1)r}$  we have

$$\dim_K \mathcal{K}^{(1)}(d) \ge n.$$

In this case the previously described "linearization" attack applies, including the breaking of the original C<sup>\*</sup> for D = 2. If n = 18 then this concerns for instance

 $d = 1 + q^4 + q^{11}$ ,

because  $d = 1 + q^{11} + q^{22} \mod (q^{18} - 1)$ .

Another remarkable special case of weak exponents occurs if condition (4) is satisfied for  $(u, v) = (\ell, k)$ , which means, assuming  $\ell < k$  w.l.o.g., that n is even and

$$r_k - r_\ell = n/2.$$

In this case the private key can be recovered from the public key. As an example, if n = 16 then  $d = 1 + q + q^5 + q^9$ , i.e.

$$\{r_0, r_1, r_2, r_3\} = \{0, 1, 5, 9\},\$$

generates such a weak system, since  $r_3 - r_1 = 8$ . The private key (up to equivalence, to be precise) could be found within a few days on a PC if we would have chosen that exponent in the above challenge.

A proper choice of the exponent d is very important and non-trivial. Our choice of d in CC11 is

$$\{r_0, r_1, r_2, r_3\} = \{0, 1, 3, 12\}.$$

Here no equality of the form (4) occurs.

### Detailed technical specification

The 25 biquadratic public polynomials  $E_i(x_0, ..., x_{24})$  (i < 25) with coefficients  $x_k$  (k < 25) in K are contained in the file Public-Key-C11.dat. It can be downloaded from the web site http://www.mystery-twister.com of the MYSTERY TWISTER competition.)

• The base field  $K = GF(2^4)$  identified with  $GF(2)^4$  via the basis

$$a^3, a^2, a, 1,$$

where a is a zero of the primitive polynomial  $x^4 + x + 1$ .

• The extension field  $L = GF(2^{100})$  of K is identified with  $K^{25}$  via the basis

$$b^{24}, b^{23}, \ldots, b, 1,$$

where b is a zero of the irreducible polynomial  $x^{25} + x + 1$ .

• According to the preceding specification we can identify

$$L = K^{25} = \mathrm{GF}(2)^{100}.$$

For the following  $Q = \sum_{i < 25} Q_i b^i \in L$  (resp.  $(Q_{24}, ..., Q_1, Q_0) \in K^{25}$ ), the unique  $P \in L$  with E(P) = Q has to be computed:

**Remark 1.** The most efficient way to repair  $C^*$  is to cancel some of the public equations  $E_i(x)$ . Concrete proposals in the NESSIE project are called SFLASH (three versions are specified, see [5]).

**Remark 2.** A draft version of CC11 with parameters m = 5 and n = 16 was broken by Dr. Faugère (private communication). It took him approximately 1/10of the effort for his mentioned breaking of the HFE Challenge 1 with n = 80and m = 1. (With parameters m = 4 and n = 20 it seems that biquadratic C<sup>\*</sup> is stronger than HFE challenge 1.)

Acknowledgement. We would like to thank Dr. Faugère for making experiments on biquadratic  $C^*$ . The results of these experiments were the base for our choice of the parameters in CC11. (Nevertheless, of course only we are responsible for this choice.)

How to submit the solution. Follow the hints on the MYSTERY TWISTER web site.

Last date for submission. December 31, 2005

**Prize.** The first person, who submits the correct solution before the end of the year 2005, will win a prize of  $5000 \in$ .

## References

- H. Dobbertin, internal reports, Bundesamt f
  ür Sicherheit in der Informationstechnik (German Federal Office for Information Security), 1993/94.
- [2] H. Dobbertin, M. Daum, P. Felke, T. Lange, G. Leander, The MYSTERY TWISTER Book, Springer-Verlag, in preparation.
- J.-C. Faugère, A. Joux, Algebraic cryptanalysis of Hidden Field Equation (HFE) cryptosystems using Gröbner bases, Advances in Cryptology - CRYPTO 2003, Lecture Notes on Computer Science, vol. 2729, Springer-Verlag, pp. 44–60.
- [4] J. Patarin, Cryptanalysis of the Matsumoto and Imai Public Key Scheme of Eurocrypt '88, Advances in Cryptology – CRYPTO '95, Lecture Notes on Computer Science, vol. 963, Springer-Verlag, pp. 248–261.
- [5] HFE web site of N. Courtois: http://www.hfe.info/

HANS DOBBERTIN and PATRICK FELKE Cryptology and IT Security Research Group (CITS) Department of Mathematics Ruhr-University of Bochum, Germany www.cits.ruhr-uni-bochum.de