Slide 1

Polynomial Equations by Re-linearization and XL

Nicolas Courtois[1,3], Alexander Klimov[2], Jacques Patarin[3], Adi Shamir[4]

[1]Paris 6 University and Toulon University, France

[2]Moscow State University, Moscow, Russia

[3]Bull CP8, Louveciennes, France

[4]The Weizmann Institute of Science, Rehovot 76100, Israel

Slide 2

Multivariate Cryptography.

(...) All the mathematicians know that
the passage from one to several variables
is an important leap that goes with complications
and requires completely new methods (...)

distinguished French mathematician

The main candidate: **HFE** (Hidden Field Equations, Eurocrypt'96).

HFE security is related to 4 difficult problems: MQ, MinRank, IP and HFE.

The present paper studies MQ - Solving **M**ultivariate **Q**uadratic Equations.

Slide 3

The problem MQ (Multivariate Quadratic)

$\mathbf{MQ(K, m, n)}$: Find a solution (at least **one**) to a system of $\mathbf{m}$ quadratic equations with $\mathbf{n}$ variables over a ring K.

$$f : \begin{cases} y_k = \sum_{i=0}^{n} \sum_{j=i}^{n} \lambda_{ijk}\ x_i x_j \\ \text{with } k = 1..m, \quad x_0 = 1 \end{cases}$$

Univariate case:

$\mathbf{MQ(\mathbb{Z}_N, 1, 1)}$ is as hard as factoring N (Rabin).

$\mathbf{MQ(GF(q), m, 1)}$ is polynomial (e.g. Berlekamp algorithm).

Multivariate case:

Fact: $\mathbf{MQ(K, m, n)}$ is NP-complete on a field, even if $K = GF(2)$.

Restriction:

From now on we consider only homogenous equations. A non-homogenous system is homogenized by adding one new variable.

Slide 4

MQ we want to solve

In cryptography we are interested in $\mathbf{MQ(K, m, n)}$ with:

- K is a small finite field $K = GF(q)$, e.g. $GF(16)$.
- The characteristic p is small, e.g. 2.
- Frequently the system is sufficiently defined or overdefined $m \geq n$.
- The system must have a solution, which is not natural for $m > n$.

Apparently, in the vast literature about solving such systems by Gröbner bases we have usually:

- K algebraically closed, thus infinite.
- The characteristic p is 0.
- The system is exactly defined $m = n$.
- The system is random.

Slide 5

Classical algorithms

Gröbner bases - an important part of applied mathematics.

The Buchberger algorithm [1965] and all the followers order the monomials in different ways and eliminate them.

The best of these algorithms we are aware of is F_5 by Jean-Charles Faugère.

The complexity of F_5:

- Proved 2^{3n} and $2^{2.7n}$ in practice.
- The exhaustive search is in q^n

Still MQ is hard for about $n > 15$.

Our discovery

The fact that the systems become much easier to solve when $m > n$ seems to have been completely overlooked so far.

Slide 6

Linearization - $m \geq n^2/2$

If $m = \varepsilon n^2$, with $\varepsilon = 1/2$ we proceed as follows:

1 Introduce new variables $y_{ij} = x_i x_j$

2 Solve the resulting linear system
(at least m equations with m variables).

$\rightarrow$ Recover the x_i from the y_{ii}.

Shamir and Kipnis in an attempt to break the HFE cryptosystem (Crypto'99) have proposed to extend linearization:

Re-linearization - $m = \varepsilon n^2$, $\varepsilon < 1/2$

3 Add also new equations such as $y_{12}y_{34} = y_{13}y_{24}$

4 We get another system to solve with $\varepsilon' > \varepsilon$.

Slide 7

Why relinearization was a bad idea

It introduces a great many additional variables.

Substitutions on those variables create more equations that can possibly be linearly independent.

The XL algorithm can be seen as an improved version of relinearization that uses only initial variables, and thus produces less unnecessary equations. It is also more flexible.

Why relinearization was a great idea

The claim [Shamir-Kipnis, Crypto'99]:

A system of εn^2 equations with n variables can be solved in expected **polynomial** time for any fixed $\varepsilon > 0$.

Result: Our experiments on XL consolidated this claim.

Slide 8

Why XL ?

E**X**tended **L**inerization or Multiply(**X**) and **L**inearize.

Conventions

$K = GF(q)$, f has m equations and n variables x_i over K.

For a given output $y \in K^m$ we put

$$l_k = f_k(x_1, \ldots, x_n) - y_k$$

The instance to solve is:

$$\begin{cases} l_1(x_1, \ldots, x_n) & = & 0 \\ & \vdots & \\ l_m(x_1, \ldots, x_n) & = & 0 \end{cases}$$

Slide 9

We consider only terms modulo the equation $a^q = a$ of the finite field K.

Powers are in the range $1, \ldots, q-1$.

Conventions, Terms

Let 1 or x^0 denote the set of constant terms.

Let x denote the set of terms $\{x_1, \ldots, x_n\}$.

Let x^k the set of all terms of degree exactly k (powers $1, \ldots, q-1$ allowed).

Conventions, Equations

We call l the set of initial equations $l_i = 0$.

We call xl the set of equations of the form $x_i l_j = 0$.

We call $x^k l$ the set of equations of the form $\prod_{j=1}^{k} x_{i_j} * l_k = 0$.

Example:

$x^2 l \cup l$ is the set of all the equations l_i and $x_i x_j l_k$ (we need $i \neq j$ if $q = 2$).

The terms present in these equations are $x^4 \cup x^2 \cup 1$.

Slide 10

Equation Sets

Let $D \in \mathbb{N}$. We call $\mathcal{I}_D$ the union of

$$\mathcal{I}_D \stackrel{def}{=} l \cup xl \cup \ldots \cup x^{D-2} l$$

Meaning of $\mathcal{I}_D$ equations:

- they are all true for the solution x.
- they are of total (multivariate) degree $\leq D$.

$\mathcal{I}_D \to \mathcal{I}_\infty$ and Vect$(\mathcal{I}_\infty = \mathcal{I}$.

$\mathcal{I}$ is the ideal spanned by the equations l_i.

The purpose

Eliminate all but **one** variable.

Theorem [extended version of this paper] relinearization technique does the same in a disguised way.

Slide 11

Description of XL

$D \in \mathbb{N}$ is the parameter of XL algorithm.

1. **Multiply:** Generate all the products $\prod_{j=1}^{k} x_{i_j} * l_i \in \mathcal{I}_D$ with $k \leq D-2$.
2. **Linearize:** Consider each monomial in x_i of degree $\leq D$ as a new variable and perform Gaussian elimination on the equations obtained in 1.

   The ordering on the monomials must be such that all the terms containing one variable (say x_1) are eliminated last.
3. **Solve:** Assume that step 2 yields at least one univariate equation in the powers of x_1. Solve this equation over the finite fields (e.g., with Berlekamp's algorithm).
4. **Repeat:** Simplify the equations and repeat the process to find the values of the other variables.

Question:

What D makes XL algorithm work for given m, n ?

Slide 12

Asymptotic analysis

Estimation of the number of equations in $\mathcal{I}_D$:

$$All \approx m \cdot n^{D-2}/(D-2)!$$

We suppose that most of them are linearly independent, $Free \approx All$.

Estimation of the number of all terms in x^D:

$$T \approx n^D/D!$$

The algorithm XL works when $Free \approx T$.

$$n^D/D! \approx mn^{D-2}/(D-2)!$$

$$n^2 \approx mD(D-1)$$

$$\boxed{D \approx \frac{n}{\sqrt{m}}}$$

Slide 13

Experiments with m=n

4 variables and 4 homogenous quadratic equations, $GF(127)$

XL equations		Δ	B	XL unknowns (B degrees)	
type	Free/All	(Free+B-T-1)		T	type
l	4/4	-6	1	10	x^2
$x^4l \cup x^2l \cup l$	122/184	-5	3	129	$x^6 \cup x^4 \cup x^2$
$x^8l \cup x^6l \cup x^4l \cup x^2l \cup l$	573/1180	-3	5	580	$x^{10} \cup x^8 \cup x^6 \cup x^4 \cup x^2$
$x^{12}l \cup x^{11}l \cup x^{10}l \cup \ldots$	3044/7280	-2	14	3059	$x^{14} \cup \ldots$
$x^{14}l \cup x^{12}l \cup x^{10}l \cup \ldots$	2677/6864	0	8	2684	$x^{16} \cup x^{14} \cup x^{12} \cup \ldots$

T: number of monomials

Δ ≥ 0 when XL solves the equations, (Δ = Free+B-T-1)

B: nb. of monomials in one variable e.g. x_1

Free/All: numbers of free/all equations of given **type**

Results

All simulations with $m = n$ showed that we need $D = 2^n$. (due to $\exists\ \bar{K}$)

Slide 14

Experiments with m=n+1

8 variables and 9 homogenous quadratic equations, $GF(127)$

XL equations		Δ	B	XL unknowns (B degrees)	
type	Free/All	(Free+B-T-1)		T	type
l	9/9	-27	1	36	x^2
$x^2l \cup l$	297/333	-68	2	366	$x^4 \cup x^2$
$x^4l \cup x^2l \cup l$	2055/3303	-25	3	2082	$x^6 \cup x^4 \cup x^2$
$x^5l \cup x^3l \cup xl$	4344/8280	-5	4	4352	$x^7 \cup x^5 \cup x^3 \cup x$
$x^6l \cup x^4l \cup x^2l \cup l$	8517/18747	3	4	8517	$x^8 \cup x^6 \cup x^4 \cup x^2$

T: number of monomials

Δ ≥ 0 when XL solves the equations, (Δ = Free+B-T-1)

B: nb. of monomials in one variable e.g. x_1

Free/All: numbers of free/all equations of given **type**

Results

All simulations with $m = n + 1$ showed that $D = n$.

Slide 15

$m = n + 2 \ldots n + 4$

8 variables and 10 homogenous quadratic equations, $GF(127)$

XL equations		Δ	B	XL unknowns (B degrees)	
type	Free/All	(Free+B-T-1)		T	type
$x^2l \cup l$	325/370	-40	2	366	$x^4 \cup x^2$
$x^3l \cup xl$	919/1280	1	3	920	$x^5 \cup x^3 \cup x$

8 variables and 12 homogenous quadratic equations, $GF(127)$

XL equations		Δ	B	XL unknowns (B degrees)	
type	Free/All	(Free+B-T-1)		T	type
xl	96/96	-31	2	128	$x^3 \cup x$
$x^2l \cup l$	366/444	1	2	366	$x^4 \cup x^2$

Slide 16

Experiments on Relinearization $D = 6$

n	m	l	n'	m"
4	8	2	9	9
4	7	3	19	19
4	6	4	34	40
4	5	5	55	86
6	10	11	363	394
6	9	12	454	548
6	8	13	559	806
6	7	14	679	1541
8	12	24	2924	3794
8	11	25	3275	4584
8	10	26	3653	5721

Number of variables in the original quadratic system

Number of equations in the original quadratic system

Number of parameters in the representation of the y_{ij}

Number of variables in the final linear system

number of equations which were required to solve the final linear system

Table 1: Experimental data for degree 6 relinearization

Slide 17

Theory

When $m \approx n$, $D \approx \frac{n}{\sqrt{m}} \approx \sqrt{n}$.

Experimental Results

- $D = 2^n$ when $m = n$.
- $D = n$ when $m = n + 1$.
- D decreases quickly for $m = n + 2$.

  $\vdots$

- It seems that indeed $D \rightarrow \mathcal{O}(\sqrt{n})$.

The simplified estimation $D \approx \frac{n}{\sqrt{m}}$ proved likely to be true when m exceeds n by a small value.

More simulations are needed.

Slide 18

The complexity of XL

Let ω be the exponent of Gaussian reduction.

$$2 \leq \omega < 3$$

For a system of $m = \varepsilon n^2$ equations with n variables we estimate:

$$D \approx \frac{n}{\sqrt{m}} \approx \lceil \frac{1}{\sqrt{\varepsilon}} \rceil$$

XL is expected to solve $m = \varepsilon n^2$ equations with n variables in **polynomial** time

$$WF \quad = \quad T^{\omega} \quad \approx \quad \mathcal{O}\left(n^{\frac{\omega}{\sqrt{\varepsilon}}}\right)$$

Slide 19

Solving systems with m close to n

- The behaviour of XL problem changes dramatically when m becomes slightly greater than n.
- q is usually small and we may guess some variables.

FXL algorithm

It is unclear how many variables should be guessed.

For example let's assume that less than $\sqrt{n}$ variables are fixed.

Then FXL is then expected to solve a system of n quadratic equations with n unknowns over $GF(q)$ in **subexponential** time:

$$WF \approx q^{\sqrt{n}} n^{\omega\sqrt{n}}$$

Slide 20

Solving MQ in Practice

FXL might be subexponential even when $m = n$.

However it becomes faster than the exhaustive search only for relatively big values of $n > 100$, for example:

Direct application to HFE Challenge 1

Acting as if they were no trapdoor in HFE.

We have $n = 80$. We expect that FXL requires a Gaussian reduction with the number of variables of about:

$$n^{\sqrt{n}}/\sqrt{n}! \approx 2^{38}$$

Current methods for sparse Gaussian elimination go up to about 2^{20} variables.

With 2^{38} variables the FXL complexity will exceed the exhaustive search in 2^{80}.

Slide 21

HFE Challenge 1

At Crypto'99 Shamir and Kipnis reduced the problem of recovering the secret key of the HFE Challenge 1 to the following problem (MinRank):

- Given n matrices M_i of size $n \times n$ over $GF(2^n)$, $n = 80$.
- Find a linear combination M of M_i that has a rank $\leq r$, $r = 7$.

Method proposed at Crypto'99

Reduce this (MinRank) problem to an overdefined instance of MQ:

- Big $K = GF(2^n)$
- $n(n-r)$ equations
- $r(n-r)+n$ variables

Slide 22

Bad news

- Solving this MQ by XL with conjectured polynomial complexity requires about 2^{152} computations.
- MinRank is NP-complete [Shallit, Frandsen, Buss 1996].
- Thus the overdefined instances of MQ generated in such a way could be harder than average and **not** be solved in polynomial time at all.

Improved method

Solve this MinRank directly in about 2^{82} [Courtois, not published yet].

Direct method

Works in 2^{62} without recovering the secret key [Courtois, not published yet].

Slide 23

Conclusion

We proved that relinearization reduces to XL algorithm.

A system of $m = \varepsilon n^2$ equations with n variables that has a solution is expected to be solved by XL in **polynomial** time of about

$$n^{\frac{\omega}{\sqrt{\varepsilon}}}$$

A system of n equations with n variables over a small finite field is expected to be solved by FXL in **subexponential** time of about

$$q^{\sqrt{n}} n^{\omega\sqrt{n}}$$

Applications in cryptography

The best known algorithms for solving multivariate equations over a very small finite field are still close to the exhaustive search.

Many cryptosystems using such equations can be proposed.

HFE can still be believed as one of the strongest candidates.