Probabilistic analysis on the rank of Macaulay matrices over finite fields

Andrea Tenti

Selmer Senter
Univerity of Bergen

Finse, May 08, 2018

Joint work with Igor Semaev

Outline

1. Algebraic attacks

2. Semiregular systems

3. Our contribution

Outline

1. Algebraic attacks

2. Semiregular systems

3. Our contribution

Algebraic Attacks

Often problems in cryptography can be reduced to solving a system of polynomial equations on a finite field. To solve such a problem, one can try to find the roots of the system. These kind of attacks are called algebraic attacks.
Some examples include:

- Find the key of AES,
- Solve Multivariate quadratic (MQ) cryptosystems,
- Decomposing a point on an elliptic curve into a sum of points with "small" coordinates, in order to perform an index calculus attack. (Summation Polynomials)

Solving such a system is considered a difficult problem.

Algebraic attacks

The most widely used methods for solving algebraic systems are XL (eXtended Linearization) and its variations, together with Gröbner-basis methods.

The methods share a common approach. Let $f_1, \dots, f_m$ be a system of polynomials in $\mathbb{F}_q$. The Macaulay matrix of degree d is computed:

$$M_d := \begin{array}{|c|c|} \hline & \text{monomials of degree } \leq d \\ \hline m_i f_j & \dots \\ \hline \end{array}$$

where m_i are monomials such that $m_i f_j$ has degree $\leq d$.

Macaulay matrix

- The choice of the monomials m_i depends on the algorithm used.
- A linear reduction is performed to find univariate polynomials or a Gröbner-basis.
- If the condition searched for by the algorithm is not found, M_{d+1} is computed and the process repeats.
- The largest degree achieved by the algorithm is called **Solving degree** (d_{solv}).
- Time-complexity is dominated by the linear algebra part of the algorithm. Hence, it depends on d_{solv} and, overall, is about $N_{d_{\mathrm{solv}}}^{\omega}$, where $2 < \omega \leq 3$ and N_d is the size of M_d.

Gröbner basis

Given a monomial order over a polynomial ring, it is possible to establish, for each polynomial f, its leading term.
A Gröbner basis of an ideal I is a set of generators G of the ideal so that $(LT(I)) = (LT(G))$.

Gröbner basis

Given a monomial order over a polynomial ring, it is possible to establish, for each polynomial f, its leading term.
A Gröbner basis of an ideal I is a set of generators G of the ideal so that $(LT(I)) = (LT(G))$.

Fact

Given a Gröbner basis G, it is possible, through a fast algorithm (FGLM, or other) to turn G into another set of generators of the form:

$$\begin{aligned}&\{p_{1,1}(x_1)\\ &\ p_{2,1}(x_1,x_2),\ldots,p_{2,t_2}(x_1,x_2),\\ &\vdots\\ &\ p_{n,1}(x_1,\ldots,x_n),\ldots,p_{n,t_n}(x_1,\ldots,x_n)\}.\end{aligned}$$

Outline

1 Algebraic attacks

2 Semiregular systems

3 Our contribution

Semiregular systems

Definition

A system $f_1, \dots, f_m \in \mathbb{F}_q[x_1, \dots, x_n]$ is called semiregular if there are no algebraic relations between the f_j of degree smaller than δ, except for the trivial ones (i.e. $f_i f_j - f_j f_i = 0$ and $f_i^q - f_i = 0$). Here, δ is the smallest degree d for which $\{LT(g) | g \in (f_1, \dots, f_m)_d\}$ is equal to the set of monomials of degree d. It is called degree of regularity.

Semiregular systems

Definition

A system $f_1, \dots, f_m \in \mathbb{F}_q[x_1, \dots, x_n]$ is called semiregular if there are no algebraic relations between the f_j of degree smaller than δ, except for the trivial ones (i.e. $f_i f_j - f_j f_i = 0$ and $f_i^q - f_i = 0$). Here, δ is the smallest degree d for which $\{LT(g) | g \in (f_1, \dots, f_m)_d\}$ is equal to the set of monomials of degree d. It is called degree of regularity.

Theorem (Bardet, Faugere, Salvy 2004)

If a system is semiregular over $\mathbb{F}_2$, the solving degree is smaller or equal than the index of the first negative coefficient of the Hilbert series

$$H_{m,n}(t) = \frac{(1+t)^n}{\Pi_{i=1}^m (1+t^{d_i})}.$$

Semiregular systems

Example (Bardet, Faugere, Salvy 2004)

For $n = m$, $q = 2$, and equations of degree D,

D	$d_{\text{solv}} \leq$
2	$0.09n + o(n)$
3	$0.15n + o(n)$
4	$0.20n + o(n)$
5	$0.24n + o(n)$
$\vdots$	$\vdots$

Semiregular systems

- For quadratic semiregular systems over $\mathbb{F}_2$, where $m \geq n^2/6$, the solving degree is ≤ 3.
- Semiregular systems are common.

Conjecture (B., F., S. 2004)

Let us consider a random system of m equations of degree $d_1, \ldots, d_m$ over $\mathbb{F}_2$ in n variables. The probability that it is semiregular tends to 1 as n increases for fixed m and d_i.

Semiregular systems

The conjecture has been proven to be false by Hodges, Molina and Schlather in 2014. Regardless, they simply believed that the formulation did not capture what exactly people meant with: ''most of the random generated systems are semiregular''.
Another conjecture was formulated:

Conjecture (Hodges, Molina, Schlather 2014)

Let $\pi(n, m, d)$ *be the proportion of systems of degree* d *with* m *polynomials in* n *variables over* $\mathbb{F}_2$ *that are semiregular. Then for every* $\epsilon > 0$

$$\pi(n, m, d) > 1 - \epsilon \qquad \textit{for every } n, m \textit{ large enough}$$

Special cases

Sometimes, polynomial systems generated by specific mathematical problems used in cryptography, behave particularly well with respect to algebraic attacks.

This means that the solving degree can be much lower than the bound stated before. Some notable examples are:

- Quadratic systems that emerge from Hidden Field Equations,
- Cubic systems that arise from summation polynomials to split points over an elliptic curve.

In both these cases, experiments show that the solving degree increases much slower (maybe it is constant) than what was predicted as the number of variables increases.

Outline

1. Algebraic attacks

2. Semiregular systems

3. Our contribution

Overdetermined systems

Let us consider a system of $m = \binom{n}{2}$ quadratic equations over $\mathbb{F}_2$.

- If the equations are linearly independent, the solving degree is 2.
- If $m \geq \binom{n}{3}/n$, then the Macaulay matrix of degree 3 is almost square.

Example

Let $f = c_{12}x_1x_2 + c_{13}x_1x_3 + c_{14}x_1x_4 + c_{23}x_2x_3 + c_{24}x_2x_4 + c_{34}x_3x_4$. The degree 3 Macaulay matrix is

$$M_3 = \begin{pmatrix} c_{23} & c_{24} & c_{34} & 0 \\ c_{13} & c_{14} & 0 & c_{34} \\ c_{12} & 0 & c_{14} & c_{24} \\ 0 & c_{12} & c_{13} & c_{23} \end{pmatrix}$$

The probability that the solving degree is bounded by 3 is 28/64.

Overdetermined systems

Problem

Given a system of polynomials in $\mathbb{F}_q[x_1, \ldots, x_n]$, find m (as a function of n) for which the probability of $d_{\text{solv}} \leq D + 1$ tends to 1, as n increases.

Overdetermined systems

Problem

Given a system of polynomials in $\mathbb{F}_q[x_1, \dots, x_n]$, find m (as a function of n) for which the probability of $d_{\text{solv}} \leq D + 1$ tends to 1, as n increases.

Theorem

Let $N := |\{$monomials of degree $D + 1$ in $\mathbb{F}_q[x_1, \dots, x_n]/(x_i^q - x_i)\}|$. If

$$m \geq \frac{N}{n},$$

then

$$\mathbb{P}(d_{\text{solv}} \leq D + 1) = 1 - (q^{N-mn} + O(nq^{-n^D})),$$

for q and D fixed and n increasing.

Overdetermined systems

- The theorem does not prove any of the conjectures formulated.
- It shows, though, that under the mentioned hypothesis, a random system behaves like a semiregular one with high probability.
- The proof of the theorem revolves around showing that the rank of the Macaulay matrix M_{D+1} is maximal.
- Analysing directly the matrix M_{D+1} is difficult.
- The trick we employed was breaking M_{D+1} in independent pieces and for every piece estimate how many vector resides in the kernel of each peace.

Current work

Goal:

- Given a random system in $\mathbb{F}_q[x_1,\ldots,x_n]$ of degree D and an integer d, understand for which m, $\mathbb{P}(d_{\mathrm{solv}} \leq D+d) \to 1$.
- We expect that

$$m \geq \frac{|\{\textit{monomials of degree } D+d \text{ in } \mathbb{F}_q[x_1,\ldots,x_n]/(x_i^q - x_i)\}|}{|\{\textit{monomials of degree } d \text{ in } \mathbb{F}_q[x_1,\ldots,x_n]/(x_i^q - x_i)\}|}$$

Conclusions

- Algebraic attacks and the problem of estimating complexity,
- Semiregular systems are system for which we can estimate the complexity and often random systems are semiregular,
- We provided a lower bound for the number of equations in order to have $d_{\mathrm{solv}} \leq D+1$.
- This bound is the one predicted for semiregular systems.