Algebraic Attacks from a Gröbner Basis Perspective

Toon Segers

October 2004

Context

- M.Sc. programme Applied Mathematics, TU/e
  Coding theory and Cryptology curriculum (CC)
- Nine-month graduation project
  Netherlands National Communications Security Agency (NLNCSA)
  General Intelligence and Security Service
- Committee
  Prof. dr. ir. H.C.A. van Tilborg, supervisor
  Dr. B.M.M. de Weger, advisor from the CC group
  Drs. G. Schmitz, advisor from the NLNCSA
  Dr. H. Sterk, advisor from the Discrete Algebra and Geometry group

Outline of the presentation

1. Motivation of this research topic
2. Mathematical background
3. Advanced techniques
4. Behavior of the techniques F4 and XL
5. XL is a Gröbner Basis algorithm
6. Analyzing XL by means of Hilbert series
7. Conclusions

1. Motivation

- Recent attention to algebraic attacks
  - 80 bit HFE challenge tackled with F5 [FJ03] and F4 [Ste04]
  - Complexity of XL could be subexponential according to [CKPS00]
- Various types of systems might be vulnerable:
  blockciphers, streamciphers and asymmetric cryptosystems
- Shannon [Sha49] states that breaking a good cipher should require:

> *"as much work as solving a system of simultaneous equations in a large number of unknowns of a complex type."*

Research interests

- The complexity of algebraic attacks, in terms of time and memory used
- A better understanding of XL
- Recent developments

Our contribution

- Simulations that illustrate the behavior
- We are getting to the bottom of XL:
  - Applying XL incrementally turns out to be a Gröbner Basis algorithm
  - Analysis based on Hilbert series might underestimate the effectiveness

2. Background

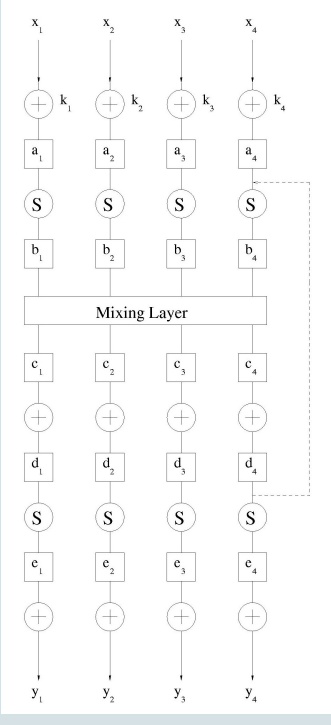
What exactly is an algebraic attack?

- Represent a cryptographic scheme as a system of equations.
- Try to solve this system.

scheme	*eqs.*	*vars.*	*field*
AES 128	$8,000$	$1,600$	$GF(2)$
AES 256	$22,400$	$4,480$	$GF(2)$
BES 128	$5,248$	$3,948$	$GF(2^8)$
HFE $n$bit	n	n	$GF(2)$

Table 2.1: Systems of quadratic equations from different schemes.

A blockcipher



The algebraic system

- For r rounds, $12 + 12r$ equations in the same number of variables
- Let $1 \leq i \leq 4$,
$$\begin{aligned} a_i &= x_i + k_i, \\ a_i b_i &= 1, \\ c_i &= \textstyle\sum_{j \neq i} b_j, \\ d_i &= c_i + k_i, \\ d_i e_i &= 1, \\ y_i &= e_i + k_i. \end{aligned}$$
- Chance of not having to invert a zero $\approx (1 - 1/256)^8 \approx 0.969$

Solving the algebraic system

- Gröbner Basis algorithms

  bring the system to the desired 'triangular form':

  $\{x_1 - h_1(x_n), \ldots, x_{n-1} - h_{n-1}(x_n), h_n(x_n)\}$

- Factorization of polynomials in one variable

  Partial factorization over small finite fields is very fast

- Measurements for a known-plaintext attack on the 32 bit blockcipher

$rounds$	$time\ (s)$	$memory\ (kb)$
1	$0,021$	400
2	$62,419$	26.261
3	$25.533,155$	1.319.550

Table 2.2: Time and memory used (Magma 2.11, 1 GHz Pentium 4)

Magma example

```
> load "fbbc.m";
Loading "fbbc.m"
> In := [[ f, f ],[ f, f ],[ f, f ],[ f, f ]];
> k := [[ 1, 3 ],[ 2, 4 ],[ 5, 7 ],[ a, c ]];
> Out := fbbc(In,k,1);
*********** 4 byte blockcipher ************

Return the ciphertext
> Out;
[[ f, 5 ], [ 5, 7 ], [ 4, 2 ], [ 3, c ]]

> F := fbbcequations(In,Out,1);
> G := GroebnerBasis(F);
> #G;
24

> G[24];
k_4^41 + (t^7 + t^6 + t^3 + 1)*k_4^40
    + (t^4 + t^3 + t^2 + 1)*k_4^39 +
    ...
    + t^6 + t^3 + t^2 + t)*k_4 + t^7 +
    t^6 + t^5 + t^4 + t^3 + t^2 + 1
```

```
> f := Factorization(G[24]);
> f;
[
    <k_4 + t^7 + t^6 + t^3 + t, 1>,
    <k_4^2 + (t^7 + t^3 + t^2 + 1)*k_4 +
        t^7 + t^5 + t^4 + t^3 + 1, 1>,
    <k_4^2 + (t^7 + t^5 + t^2 + t + 1)*k_4 +
        t^6 + t^4 + t^3 + t^2, 1>,
    <k_4^16 + (t^3 + t^2 + t)*k_4^15 +
        (t^7 + t^5 + t^2 + 1)*k_4^14 + (t^7 + t^6
        ...
        t^2, 1>,
    <k_4^20 + (t^5 + t^2 + t + 1)*k_4^19 +
        (t^6 + t^5 + 1)*k_4^18 + (t^7 + t^4 +
        ...
        + 1)*k_4^2 + (t^7 + t^5 + t^4 + 1)*k_4 +
        t^6 + t^5 + t^4 + t^3 + t^2, 1>
]

> a := Evaluate(f[1,1],24,0);
> a;
t^7 + t^6 + t^3 + t

> Evaluate(G[23],24,a);
k_3 + t^6 + t^5 + t^4 + t^2 + 1
```

Systems of equations and ideals

- Every equation describing the scheme corresponds to a polynomial.
- A solution to the equation is a root of the polynomial.
- The algebraic system is represented by a set of polynomials F.

$$\begin{array}{ll} a_i & = x_i + k_i, \\ a_i b_i & = 1, \\ c_i & = \sum_{j \neq i} b_j, \\ d_i & = c_i + k_i, \\ d_i e_i & = 1, \\ y_i & = e_i + k_i. \end{array} \quad \rightarrow \quad F = \left\{ \begin{array}{l} a_i - x_i - k_i, \\ a_i b_i - 1, \\ c_i - \sum_{j \neq i} b_j, \\ d_i - c_i - k_i, \\ d_i e_i - 1, \\ y_i - e_i - k_i \end{array} \right\}$$

Systems of equations and ideals (2)

Definition 1. Let $P = k[x_1, \ldots, x_n]$ be a polynomial ring.
A subset $I \subset P$ is an *ideal* if it satisfies:

1. $0 \in I$;
2. If $f, g \in I$, then $f + g \in I$;
3. If $f \in I$ and $h \in P$, then $hf \in I$.

Solutions to the algebraic system correspond to common zeros of the elements in the ideal spanned by F.

Example 2.

$$F = \{x, y^2\} \subset k[x, y] \quad \rightarrow \quad I = \{x, y^2, x + y^2, xy, \ldots\}$$

Gröbner Bases

Definition 3. $LT(f)$ and $LM(f)$ are called the *leading term* and *leading monomial* of the polynomial f (with respect to a predetermined ordering).

Example 4.

$$LT(5x^2 + xy + 7y^3) = 5x^2, \quad LM(5x^2 + xy + 7y^3) = x^2$$

Definition 5. A finite subset $G = \{g_1, \ldots, g_m\}$ of the ideal $I \subseteq k[x_1, \ldots, x_n]$ is called a *Gröbner Basis* if

$$\langle LT(g_1), \ldots, LT(g_m) \rangle = \{LT(f) : f \in I\}.$$

Example 6.

$$F = \{x + y + z, xy + yz, y + z\} \qquad \subset GF(2)[x, y, z] \quad (lex)$$
$$G = \{x + y + z, xy + yz, y + z, z^2\}$$

The S-polynomial

Definition 7. Let f and g be two polynomials in $k[x_1, \ldots, x_n]$.
Define x^γ as

$$x^\gamma = LCM(LM(f), LM(g)).$$

The *S-polynomial* of the pair (f, g) is equal to the sum

$$S(f, g) = \frac{x^\gamma}{LT(f)} \cdot f - \frac{x^\gamma}{LT(g)} \cdot g.$$

The pair (f, g) that is used to form the S-polynomial is commonly referred to as *critical pair*.

Example 8. *The S-polynomial corresponding to the critical pair* $(x + y + z, xy + yz)$ *equals*

$$y(x + y + z) - (xy + yz) = y^2.$$

The Buchberger Algorithm

- Input: $F = \{f_1, \ldots, f_{m'}\}$
- Output: A Gröbner Basis $G = \{g_1, \ldots, g_m\}$ of I where $F \subset G$.

1. $G := F$
2. $REPEAT$
3. $\quad G' := G$
4. $\quad FOR$ each pair $(p, q), p \neq q$ in G' DO
5. $\quad\quad S =$ the remainder of $S(p, q)$ after division by elements of G'
6. $\quad\quad IF\ S \neq 0\ THEN\ G := G \cup \{S\}$
7. $UNTIL\ G = G'$

Example of the Buchberger Algorithm

1. Let $F = \{x + y + z, xy + yz, y + z\} \subset GF(2)[x, y, z]$ (lex).
2. The S-polynomial of the pair (f_1, f_2) is $y(x + y + z) - (xy + yz) = y^2$.
3. The remainder of y^2 after division by F is

$$y^2 - (y + z)f_3 = z^2.$$

   Append z^2 to the current basis.

4. Calculate the S-polynomial corresponding to the pair (f_1, f_3),

$$y(x + y + z) - x(y + z) = xz + y^2 + yz.$$

   This reduces to zero after division, just like every other critical pair.

5. It follows that the set $\{x + y + z, xy + yz, y + z, z^2\}$ is a Gröbner Basis of the ideal spanned by F.

3. Advanced techniques

- Strategies and criteria
  - Strategy: involves the selection and reduction of critical pairs
  - Criterion: determines which pairs are not taken into account
  - Examples:
    F4, Gebauer and Möller Installation [GM88] and F5 [Fau02]
- F4 by Faugère [Fau99]
- XL by Courtois, Klimov, Patarin and Shamir [CKPS00]
- F5 by Faugère [Fau02]

The algorithm F4

- Application of linear algebra is a generalization of the Sylvester matrix for the computation of resultants [Laz83].
- F4 is in fact an implementation of a strategy.
- Select a set of critical pairs, for example of the same degree.
- Represent polynomials as rows in a matrix.
- Reduce these polynomials simultaneously by Gaussian elimination.
- It is possible to apply criteria, like the Gebauer and Möller Installation.

$$\begin{aligned} f_1 &= x^2 + y^2 \\ f_2 &= x^2 + 1 \\ f_3 &= x^2 + x \\ f_4 &= y^2 + y \end{aligned} \rightarrow \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \rightarrow \begin{aligned} g_1 &= x^2 + 1 \\ g_2 &= x + 1 \\ g_3 &= y^2 + 1 \\ g_4 &= y + 1 \end{aligned}$$

Extended Linearization, XL

$F = \{f_1, \ldots, f_m\} \subset k[x_1, \ldots, x_n]$

Multiply Create the following set of monomial multiplications: $\left\{x^\alpha f_i : \alpha \in \mathbb{Z}^n_{\geq 0}, i \in \{1, \ldots, m\}, deg(x^\alpha f_i) \leq D\right\}$

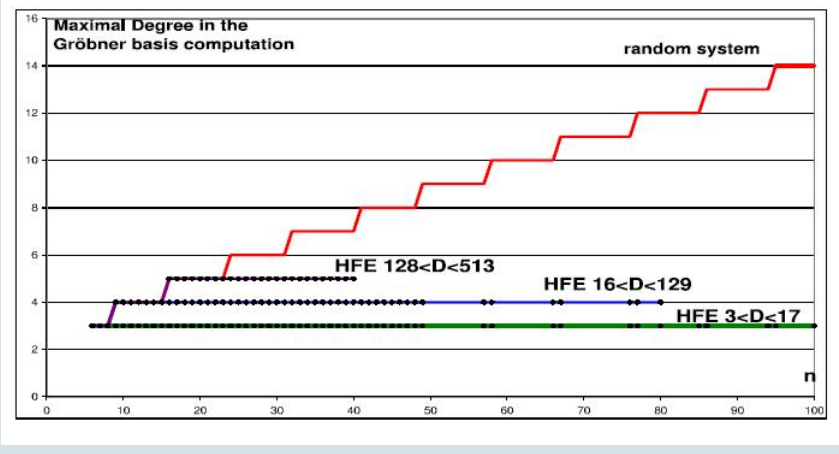
Linearize Consider each monomial x^α, $\alpha \in \mathbb{Z}^n_{\geq 0}$, of degree $\leq D$ as a new variable and perform Gaussian elimination on the equations obtained in Step 1. The ordering on the monomials must be such that all the terms containing one variable, say x_n, are eliminated last;

Solve If D is chosen such that Step 2 yields at least one univariate equation in the powers of x_n, then solve this equation over the finite field k, for example with Berlekamp's algorithm;

Repeat Simplify the equations with the solutions from Step 3 and repeat the process to find the values of the remaining variables.

The algorithm F5 applied to HFE

For all HFE [Pat96] instances with a secret HFE polynomial having fixed degree, the degrees of the intermediate polynomials during the computation with F5 are bounded.



4. Behavior of the techniques F4 and XL

- XL creates larger matrices than F4.
- The complexity of the Gaussian elimination step is crucial.

$$\mathcal{O}(\binom{n+D}{D}^{\omega}) \approx \mathcal{O}((\frac{n^D}{D!})^{\omega}) \text{ where } \omega = 2.376.$$

- XL consumes more time and memory than F4.
- Applying XL incrementally is a Gröbner Basis algorithm that is very similar to F4.
- XL returns a Gröbner Basis while keeping the degree of the intermediate polynomials lower during computation.

The algorithm F4

The implementation of F4 in Magma 2.11 applied to a random system of multivariate, quadratic equations:

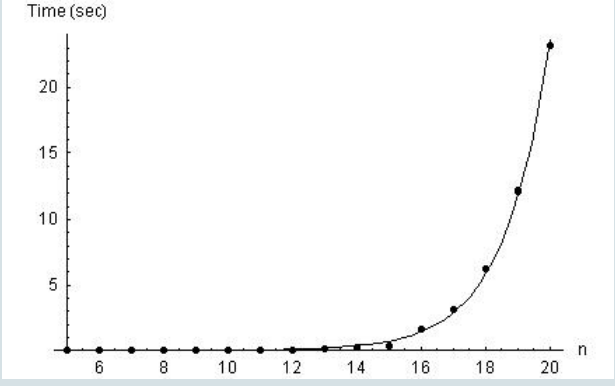
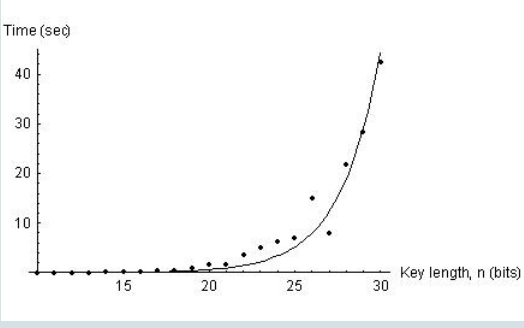
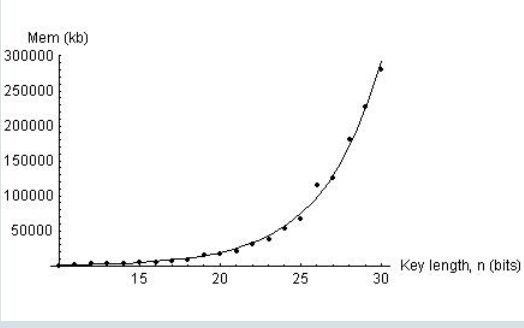


Figure 4.1: Exponential fit $0.0000193 * 2.02^n$.

F4 applied to HFE

Average time and memory consumption for an attack on HFE with the F4 implementation of Magma. The complexity seems to be exponential.





The exponential fits are $0.000118 * 1.53^n$ and $78.6 * 1.32^n$ respectively.

Size of the matrices during XL and F4

XL matrices have roughly twice as many rows as those that occur during F4.

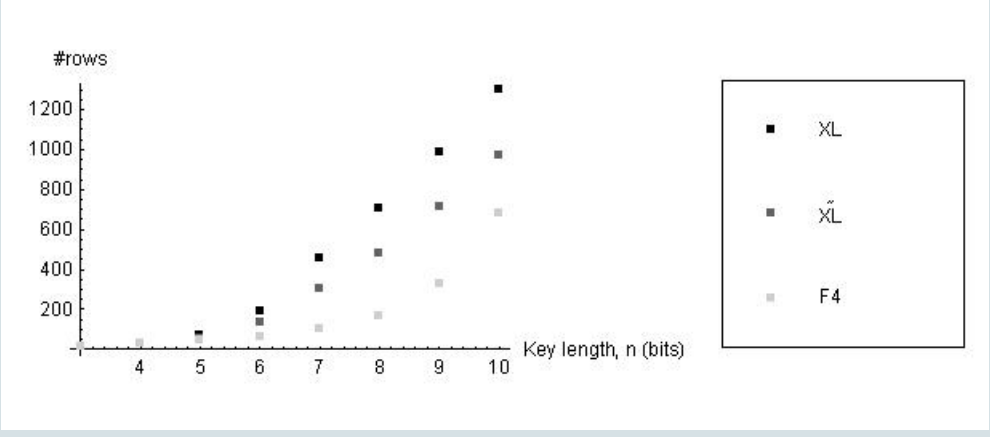


Figure 4.2: $\widetilde{XL}$ is the XL matrix after Gaussian elimination

Maximum degree during the algorithms

d_{F4} The maximum degree of a polynomial during the F4 algorithm

d_{XL} The minimum degree at which XL finds a Gröbner Basis

d_{min} The minimum degree at which XL finds a partial solution

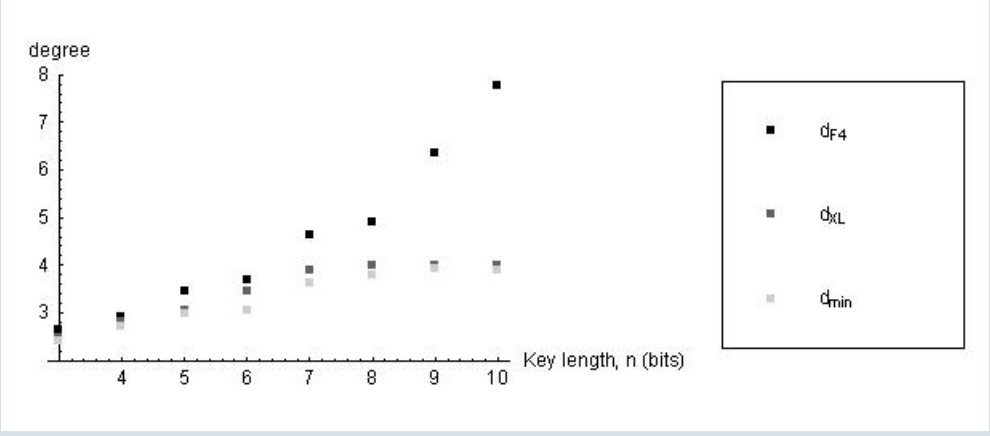


Figure 4.3: XL and F4 applied to HFE

5. XL is a Gröbner Basis algorithm

- Result appeared in May 2004 on the IACR e-print server [SKI04].
- F4 selects pairs (b_1, b_2) and reduces the matrix corresponding to the set of polynomials

$$L_d = \bigcup_{(b_1,b_2)\in B_d} \left\{ \frac{LCM(LM(b_1), LM(b_2))}{LT(b_1)} b_1, \frac{LCM(LM(b_1), LM(b_2))}{LT(b_2)} b_2 \right\}.$$

- XL selects all possible critical pairs (b_1, b_2) and creates the set

$$XL_d = \bigcup_{(b_1,b_2)\in B_d, i=1,2} \left\{ x^\alpha b_i : \alpha \in \mathbb{Z}^n_{\geq 0}, deg(x^\alpha b_i) \leq d \right\}.$$

- The matrices during F4 are parts of matrices that occur during the repeated application of XL.

6. Analysis of XL based on Hilbert series

- Original idea presented by Moh in 2001 [Moh01].
- Diem will give more theoretical background on Asiacrypt 2004 [Die04].
- Hilbert series give the dimension of the vector space corresponding to the elements of a given degree of a homogeneous ideal.
- The vector space spanned by the polynomials of XL for degree D is isomorphic to the elements of degree D in the ideal spanned by the homogenized original polynomials.
- An ideal spanned by homogeneous polynomials is homogeneous.
- If the dimension of this 'XL vector space' is close to the dimension of the polynomial ring, it is likely that one is able to eliminate a variable.

Analysis of XL based on Hilbert series (2)

- Let $F = \{f_1, \ldots, f_m\} \subset P = k[x_1, \ldots, x_n]$ be the original set of polynomials.
- The ideal I is spanned by $\{f_1^h, \ldots, f_m^h\} \subset k[x_0, x_1, \ldots, x_n]$.
- Hilbert forms give the vector space dimension of the quotient space $k[x_0, x_1, \ldots, x_n]_D / I_D$, where I is a homogeneous ideal:

$$\begin{aligned} HF_I(D) &= dim_k(k[x_0, x_1, \ldots, x_n]_D / I_D) \\ &= dim_k(k[x_0, x_1, \ldots, x_n]_D) - dim_k(I_D). \end{aligned}$$

- Diem defines $U_{\leq D}$ as the vector space spanned by

$$\left\{ x^\alpha f_i : \alpha \in \mathbb{Z}_{\geq 0}^n, 1 \leq i \leq m, deg(x^\alpha f_i) \leq D \right\}.$$

Analysis of XL based on Hilbert series (3)

- Diem and Moh define

$$\chi(D) = dim_k(P_{\leq D}) - dim_k(U_{\leq D}).$$

- If all monomials occur in the support of the ideal and $\chi(D) \leq D + 1$, then there is a univariate polynomial in the basis of the vector space spanned by $U_{\leq D}$.
- However, this is not a necessary condition to find a partial solution.
- Hilbert forms are the coefficients of Hilbert series:

$$HS(I) = \sum_{D=1}^{\infty} HF_I(D)T^D \in \mathbb{Z}[[T]].$$

Analysis of XL based on Hilbert series (4)

- $U_{\leq D}$ is the vector space spanned by

$$\left\{x^\alpha f_i : \alpha \in \mathbb{Z}^n_{\geq 0}, 1 \leq i \leq m, deg(x^\alpha f_i) \leq D\right\}.$$

- This space is isomorphic to the vector space spanned by

$$\left\{x^\alpha f_i^h : \alpha \in \mathbb{Z}^{n+1}_{\geq 0}, 1 \leq i \leq m, deg(x^\alpha f_i^h) = D\right\}.$$

- This is equal to the elements of degree D in the ideal I, therefore

$$\begin{aligned}\chi(D) &= dim_k(P_{\leq D}) - dim_k(U_{\leq D}) \\ &= dim_k(k[x_1, \ldots, x_n]_{\leq D}/U_{\leq D}) \\ &= dim_k(k[x_0, x_1, \ldots, x_n]_D/I_D) \\ &= HF_I(D).\end{aligned}$$

Analysis of XL based on Hilbert series (5)

- Hilbert series of generic systems give a lower bound for $HF_I(D)$ under the Maximal Rank Conjecture [Val96]

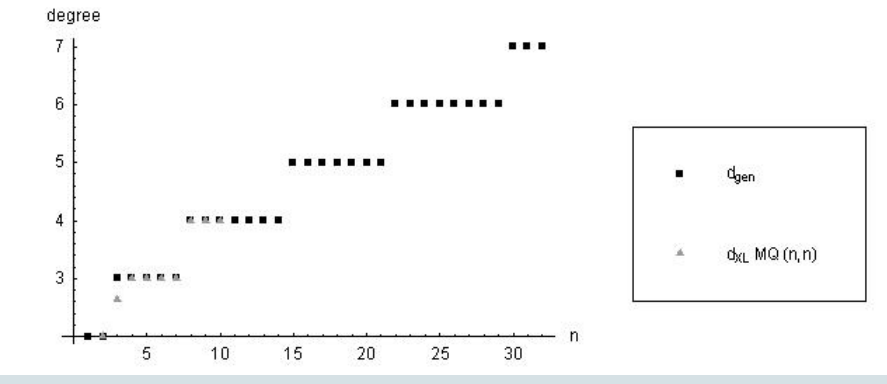


Figure 6.1: $HS_{gen} = \frac{(1+T)^{n+1}}{\prod_{j=1}^{m}(1+T^2)}$

7. Conclusions

- F4 is faster and more efficient than XL, unless d_{XL} is bounded for increasing n as we have seen for HFE.
- XL is a tool to distinguish between cryptosystems with more structure like HFE and random algebraic systems, on the basis of a smaller d_{XL}.
- The robustness of cryptosystems against XL might be overestimated, since the condition $\chi(D) \leq D + 1$ is not often necessary.

References

[CKPS00] N. Courtois, A. Klimov, J. Patarin, and A. Shamir, *Efficient Algorithms for Solving Overdefined Systems of Multivariate Polynomial Equations*, Advances in Cryptology - Eurocrypt 2000 **1807** (2000), 392–407.

[FJ03] J.-C. Faugère and A. Joux, *Algebraic Cryptanalysis of Hidden Field Equation (HFE) Cryptosystems Using Gröbner Bases*, Advances in Cryptology - Crypto 2003 **2729** (2003), 44–60.

[Ste04] A. Steel, *Allan Steel's Gröbner Basis timings page*, http://magma.maths.usyd.edu.au/users/allan/gb/, May 2004.

[Sha49] C.E. Shannon, *Communication Theory of Secrecy Systems*, Bell System Technical Journal (1949), no. 28.

References

[Fau99] J.-C. Faugère, *A New Efficient Algorithm for Computing Gröbner Bases (F4)*, Journal of Pure and Applied Algebra **139** (1999), 61–88.

[Fau02] J.-C. Faugère, *A New Efficient Algorithm for Computing Gröbner Basis without Reduction to Zero (F5)*, Proceedings of the 2002 international symposium on Symbolic and algebraic computation, ACM Special Interest Group on Symbolic and Algebraic Manipulation, 2002.

[GM88] R. Gebauer and H.M. Möller, *On an Installation of Buchberger's Algorithm*, Journal of Symbolic Computation **6** (1988), 275–286.

[Laz83] D. Lazard, *Gröbner Bases, Gaussian Elimination and Resolution of Systems of Algebraic Equations*, Proc. Eurocal '83 **162** (1983), 146–157.

References

[SKI04] M. Sugita, M. Kawazoe, and H. Imai, *Relation between XL algorithm and Gröbner Bases Algorithms*, IACR eprint server, http://eprint.iacr.org/2004/112/, May 2004.

[Moh01] T. Moh, *On the Method of "XL" and its Inefficiency to TTM*, IACR eprint server, http://eprint.iacr.org/2001/047/, June 2001.

[Die04] C. Diem, *An analysis of the XL-algorithm*, Private communication, April 2004. (Accepted for Asiacrypt 2004)

[Val96] G. Valla, *Six Lectures on Commutative Algebra*, Progress in Mathematics, vol. 166, ch. Problems and Results on Hilbert Polynomials of Graded Algebras, Birkhäuser Verlag, Basel, 1996.

References

[Pat96] J. Patarin, *Hidden Field Equations (HFE) and Isomorphism of Polynomials (IP): Two New Families of Asymmetric Algorithms*, Advances in Cryptology - Eurocrypt '96 **1070** (1996), 33–48.

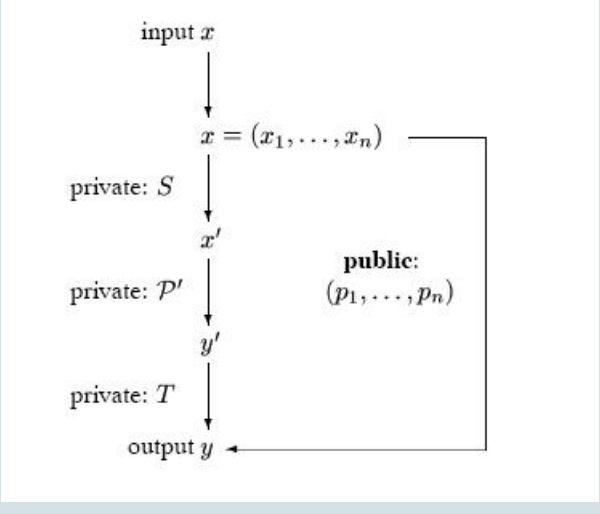


Figure 7.1: A schematic description of HFE from (PW04).