Algebraïsche Aanvallen met behulp van Gröbner Bases

Toon Segers

Oktober 2004

Context

- MSc. programma Technische Wiskunde, TU/e
  Coderingstheorie en cryptologie vakgroep (CC)
- Afstudeerproject van 9 maanden
  Nationaal Bureau voor Verbindingsbeveiliging (NBV)
  Algemene Inlichtingen- en Veiligheidsdienst
- Afstudeercommissie
  Prof. dr. ir. H.C.A. van Tilborg, leerstoelhouder CC
  Dr. B.M.M. de Weger, begeleider vakgroep CC
  Drs. G. Schmitz, begeleider NBV
  Dr. H. Sterk, lid vakgroep Discrete Algebra en Meetkunde

Opzet

1. Aanleiding voor dit onderzoek
2. Wiskundige achtergrond
3. Geavanceerde technieken
4. Gedrag van de methoden F4 en XL
5. XL is een Gröbner Basis algoritme
6. Analyse van XL met behulp van Hilbertrijen
7. Conclusie

1. Aanleiding

- Recente aandacht voor algebraïsche aanvallen
  - 80 bit HFE challenge gebroken met F5 [FJ03] en F4 [Ste04]
  - XL gedraagt zich subexponentieel volgens [CKPS00]
- Allerlei typen systemen zijn mogelijk kwetsbaar:
  blockciphers, streamciphers en asymmetrische systemen
- Shannon [Sha49] zegt over de werklast van een aanval:

> *"[it should take] as much work as solving a system of simultaneous equations in a large number of unknowns of a complex type."*

Nieuwsgierig naar

- De complexiteit van de algebraïsche aanval, uitgedrukt in tijd en geheugengebruik
- Een beter begrip van XL
- Recente ontwikkelingen

Onze bijdrage

- Simulaties die het gedrag illustreren
- Het doorgronden van XL:
  - Herhaald toepassen van XL blijkt een Gröbner Basis algoritme
  - Analyse van XL op basis van Hilbertrijen onderschat de effectiviteit

2. Achtergrond

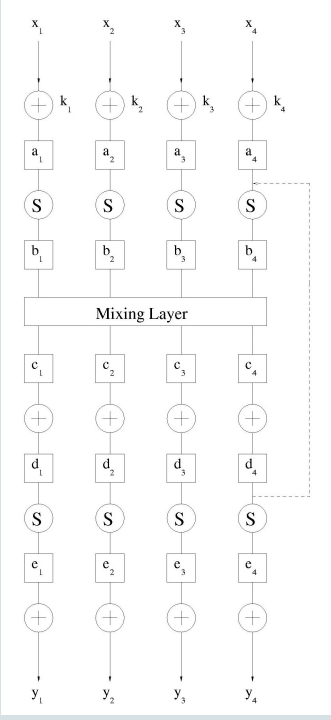
Wat is een algebraïsche aanval?

- Stel een cryptografisch systeem voor als een stelsel vergelijkingen.
- Los dit stelsel op.

systeem	*vgl.*	*var.*	*lichaam*
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	$GF(2)$

Tabel 2.1: Kwadratische stelsels vergelijkingen van verschillende systemen.

Een blockcipher



De vergelijkingen

- Voor r ronden, $12 + 12r$ vergelijkingen in evenzoveel variabelen
- Zij $1 \leq i \leq 4$,
$$\begin{aligned} a_i &= x_i + k_i, \\ a_i b_i &= 1, \\ c_i &= \textstyle\sum_{j \neq i} b_j, \\ d_i &= c_i + k_i, \\ d_i e_i &= 1, \\ y_i &= e_i + k_i. \end{aligned}$$
- Kans op geen 'nul inverse' $\approx (1 - 1/256)^8 \approx 0.969$

Het oplossen

- Gröbner Basis algoritmen,

  brengen het multivariate stelsel in gewenste 'driehoeksvorm':

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

- Factoriseren van polynomen in een variabele

  Partieel factoriseren over kleine eindige lichamen is snel

- Metingen voor een known-plaintext aanval op het 32 bit blockcipher

$ronden$	$tijd\ (s)$	$geheugen\ (kb)$
1	$0,021$	400
2	$62,419$	26.261
3	$25.533,155$	$1.319.550$

Tabel 2.2: Tijd en geheugengebruik, Magma 2.11 (1 GHz Pentium 4)

Voorbeeld in Magma

```
> 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
```

Stelsels vergelijkingen en idealen

- Iedere vergelijking in het stelsel wordt voorgesteld als een polynoom.
- Een oplossing van de vergelijking is een nulpunt van het polynoom.
- Een stelsel wordt nu voorgesteld als een verzameling polynomen F.

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

Stelsels vergelijkingen en idealen (2)

Definitie 1. Zij $P = k[x_1, \ldots, x_n]$ een polynoomring.
Een deelverzameling $I \subset P$ is een *ideaal* als het voldoet aan:

1. $0 \in I$;
2. Als $f, g \in I$, dan $f + g \in I$;
3. Als $f \in I$ en $h \in P$, dan $hf \in I$.

Oplossingen van het stelsel corresponderen met gemeenschappelijke nulpunten van elementen in het ideaal voortgebracht door F.

Voorbeeld 2.

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

Gröbner Bases

Definitie 3. $LT(f)$ en $LM(f)$ zijn de *kopterm* en het *kopmonoom* van het polynoom f (ten opzichte van een vooraf bepaalde ordening).

Voorbeeld 4.

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

Definitie 5. Een eindige deelverzameling $G = \{g_1, \ldots, g_m\}$ van het ideaal $I \subseteq k[x_1, \ldots, x_n]$ is een *Gröbner Basis* als

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

Voorbeeld 6.

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

Het S-polynoom

Definitie 7. Zij f en g twee polynomen in $k[x_1, \ldots, x_n]$.
Definieer x^γ als

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

Het *S-polynoom* van het paar (f, g) is de combinatie

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

Het paar (f, g) dat gebruikt wordt om een S-polynoom te vormen wordt meestal een *kritisch paar* genoemd.

Voorbeeld 8. *Het S-polynoom bij het kritische paar* $(x + y + z, xy + yz)$ *is*

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

Het Buchberger Algoritme

- Argument: $F = \{f_1, \ldots, f_{m'}\}$
- Uitkomst: Een Gröbner Basis $G = \{g_1, \ldots, g_m\}$ voor I met $F \subset G$.

1. $G := F$
2. $REPEAT$
3. $\quad G' := G$
4. $\quad FOR$ ieder paar $(p, q), p \neq q$ in G' DO
5. $\quad\quad S =$ de rest na deling van $S(p, q)$ door elementen van G'
6. $\quad\quad IF\ S \neq 0\ THEN\ G := G \cup \{S\}$
7. $UNTIL\ G = G'$

Voorbeeld van het Buchberger Algoritme

1. Zij $F = \{x + y + z, xy + yz, y + z\} \subset GF(2)[x, y, z]$ (lex).
2. Het S-polynoom bij het paar (f_1, f_2) is $y(x + y + z) - (xy + yz) = y^2$.
3. De rest van y^2 na staartdeling met F is

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

   Voeg z^2 aan de basis toe.

4. Bereken het S-polynoom bij paar (f_1, f_3),

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

   Deze reduceert tot nul na deling met rest, net als alle andere paren.

5. Hieruit volgt dat de verzameling $\{x + y + z, xy + yz, y + z, z^2\}$ een Gröbner Basis is van het ideaal opgespannen door F.

3. Geavanceerde technieken

- Strategieën en criteria
  - Strategie: behelst de selectie en reductie van paren
  - Criterium: bepaalt welke paren niet meegenomen hoeven te worden
  - Voorbeelden:

    F4, Gebauer en Möller Installatie [GM88] en F5 [Fau02]
- F4 door Faugère [Fau99]
- XL door Courtois, Klimov, Patarin en Shamir [CKPS00]
- F5 door Faugère [Fau02], de maximale graad van tussentijdse polynomen tijdens Faugère's implementatie van F5 is begrensd wanneer deze wordt toegepast op HFE voor een geheim polynoom met vaste graad

Het algoritme F4

- Toepassing van lineaire algebra,
  generalisatie van Sylvester matrix om resultante te berekenen [Laz83]
- F4 is een implementatie van een strategie, ruimte voor criterium
- Selecteer een groep paren, bijvoorbeeld van dezelfde graad
- Stel de polynomen voor als rijen in een matrix
- Pas simultane reductie toe door Gauss eliminatie
- Pas eventueel criteria toe, zoals de Gebauer en Möller Installatie

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

Extended Linearization, XL

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

Multiply Maak de volgende verzameling van vermenigvuldigingen:
$\left\{x^\alpha f_i : \alpha \in \mathbb{Z}^n_{\geq 0}, i \in \{1, \ldots, m\}, deg(x^\alpha f_i) \leq D\right\}$

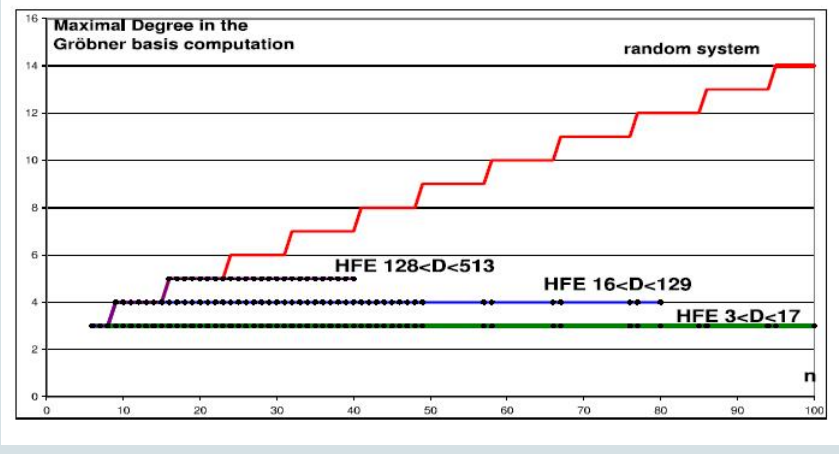
Linearize Beschouw ieder monoom x^α, $\alpha \in \mathbb{Z}^n_{\geq 0}$, van graad $\leq D$ als een nieuwe variabele en pas Gauss eliminatie toe op het stelsel van Stap 1. De ordering dient zodanig te zijn dat een variabele, zeg x_n, als laatste geëlimineerd wordt;

Solve Als D zodanig is gekozen dat Stap 2 ten minste een univariaat polynoom oplevert, los dan deze vergelijking op;

Repeat Vul de gevonden oplossing van Stap 3 in het nieuwe stelsel in en herhaal het proces.

Het algoritme F5 toegepast op HFE

Voor alle HFE [Pat96] instanties waarbij het geheime HFE polynoom een vaste graad heeft, zijn de graden van de polynomen tijdens de berekening met F5 begrensd.



4. Gedrag van de methoden F4 en XL

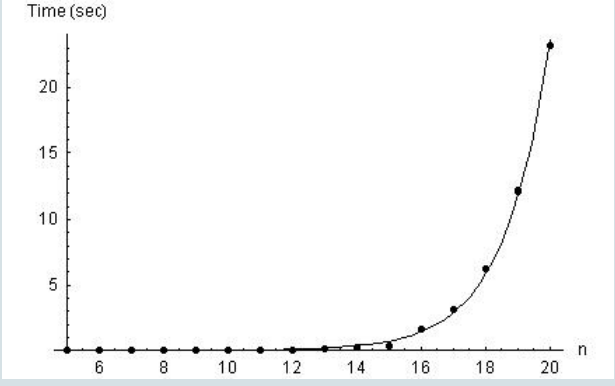
- Toepassen van XL leidt tot grotere matrices dan F4.
- De complexiteit van Gauss eliminatie is bepalend.

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

- XL gebruikt meer tijd en geheugen dan F4.
- Herhaald toepassen van XL is in feite een Gröbner Basis algoritme dat lijkt op F4.
- XL vindt een Gröbner Basis voor een lagere graad van de tussentijdse polynomen.

Het algoritme F4

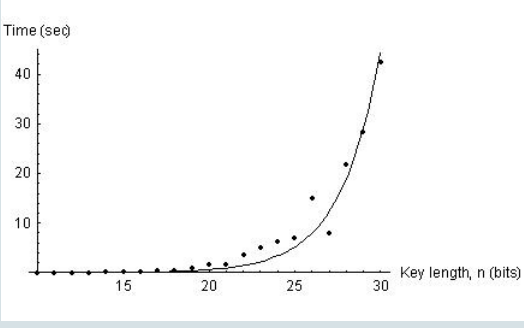
De F4 implementatie van Magma 2.11 toegepast op een willekeurig stelsel kwadratische vergelijkingen:

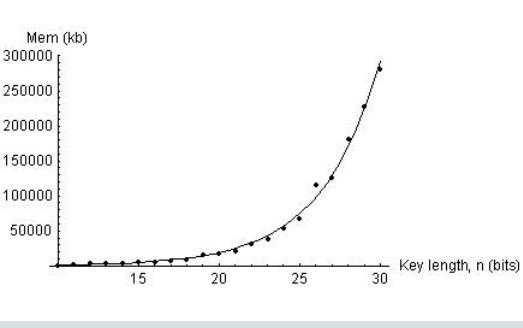


Figuur 4.1: Exponentiële fit $0.0000193 * 2.02^n$.

F4 toegepast op HFE

Gemiddelde tijd en geheugengebruik voor een aanval op HFE met Magma's F4. Het verloop lijkt exponentieel.

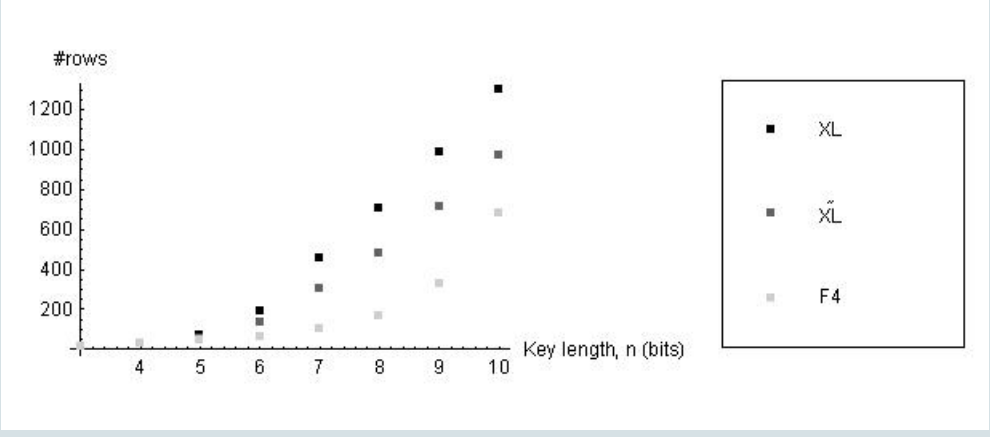




De exponentiële fits zijn respectievelijk $0.000118 * 1.53^n$ en $78.6 * 1.32^n$.

Grootte van de matrices bij XL en F4

De XL matrices hebben ruim twee keer zoveel rijen als de grootste bij F4.



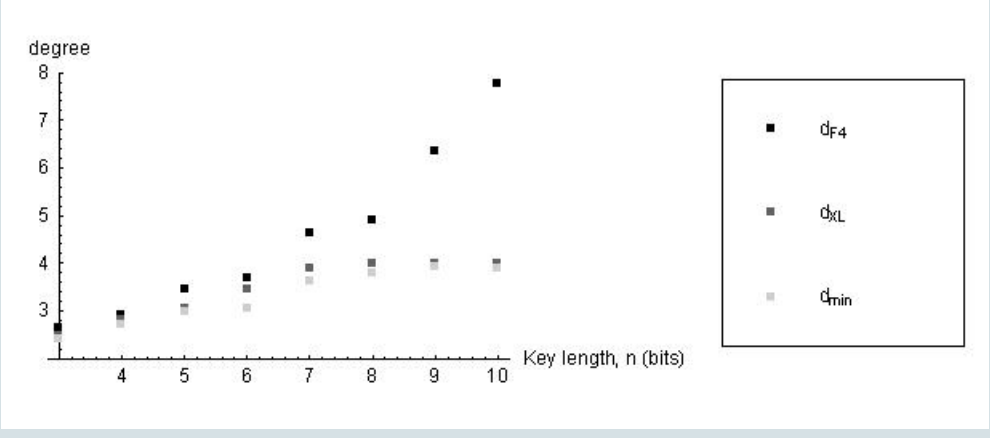
Figuur 4.2: $\widetilde{XL}$ is de XL matrix na Gauss eliminatie

Maximale graad van tussentijdse berekeningen

d_{F4} De maximale graad van polynomen tijdens het F4 algoritme

d_{XL} De minimale graad waarbij XL een Gröbner Basis vindt

d_{min} De minimale graad waarvoor XL een partiële oplossing vindt



Figuur 4.3: XL en F4 toegepast op HFE

5. XL is een Gröbner Basis algoritme

- Resultaat verscheen in mei 2004 op de IACR e-print server [SKI04].
- F4 selecteert paren (b_1, b_2) en reduceert de matrix die correspondeert met de verzameling polynomen

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

- XL selecteert alle mogelijke paren (b_1, b_2) en creëert de verzameling

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

- De matrices tijdens F4 zijn delen van de matrices die voorkomen bij het herhaald toepassen van XL.

6. Analyse van XL met Hilbertrijen

- Oorspronkelijke idee van Moh in 2001 [Moh01].
- Diem geeft meer achtergrond op Asiacrypt 2004 [Die04].
- Hilbertrijen geven de dimensie van de vectorruimte die correspondeert met elementen van een homogeen ideaal tot een bepaalde graad.
- De vectorruimte opgespannen door de polynomen van XL voor graad D is isomorf met de elementen van graad D in het ideaal opgespannen door de gehomogeniseerde oorspronkelijke polynomen.
- Een ideaal opgespannen door homogene polynomen is homogeen.
- Als de dimensie van die 'XL vectorruimte' dicht in de buurt ligt van de vectorruimte dimensie van de polynoomring, dan kan men waarschijnlijk een variabele elimineren.

Analyse van XL met Hilbertrijen (2)

- Zij $F = \{f_1, \ldots, f_m\} \subset P = k[x_1, \ldots, x_n]$ het oorspronkelijke stelsel.
- Het ideaal I wordt opgespannen door $\{f_1^h, \ldots, f_m^h\} \subset k[x_0, x_1, \ldots, x_n]$.
- Hilbert forms geven de vectorruimte dimensie van de quotiëntruimte $k[x_0, x_1, \ldots, x_n]_D / I_D$, waarbij I een homogeen ideaal is:

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

- Diem definieert $U_{\leq D}$ als de vectorruimte opgespannen door

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

Analyse van XL met Hilbertrijen (3)

- Diem en Moh stellen beide

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

- Als alle monomen voorkomen in het ideaal en $\chi(D) \leq D + 1$, dan is er een univariaat polynoom in de basis van de vectorruimte $U_{\leq D}$.
- Dit is echter geen noodzakelijke voorwaarde voor het vinden van een partiële oplossing.
- Hilbertrijen hebben Hilbert forms als coëfficiënten:

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

Analyse van XL met Hilbertrijen (4)

- $U_{\leq D}$ is de vectorruimte opgespannen door

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

- Deze ruimte is isomorf met de vectorruimte opgespannen door

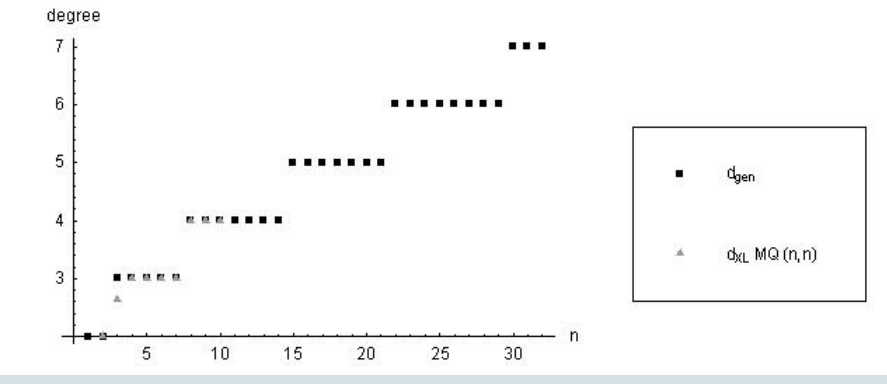
$$\left\{x^\alpha f_i^h : \alpha \in \mathbb{Z}^{n+1}_{\geq 0}, 1 \leq i \leq m, deg(x^\alpha f_i^h) = D\right\}.$$

- Dit is gelijk aan de elementen van graad D in het ideaal I, dus

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

Analyse van XL met Hilbertrijen (5)

- Hilbertrijen van generieke stelsels geven een ondergrens voor $HF_I(D)$ onder de Maximal Rank Conjecture [Val96]



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

7. Conclusies

- F4 is sneller en efficiënter dan XL
  tenzij d_{XL} constant blijft voor grotere n, zoals bij HFE.
- XL is een methode om zwakke cryptosystemen te onderscheiden
  op basis van lagere d_{XL}.
- De bestendigheid van cryptosystemen tegen XL wordt overschat
  doordat $\chi(D) \leq D + 1$ vaak geen noodzakelijke voorwaarde is.

Literatuur

[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.

Literatuur

[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.

Literatuur

[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.

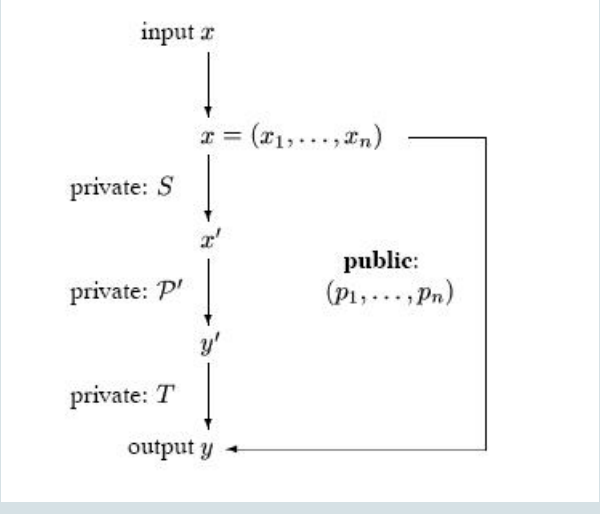
[Moh01] T. Moh, *On the Method of "XL" and its Inefficiency to TTM*, IACR eprint server, http://eprint.iacr.org/2001/047/, June 2001.

[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.

Literatuur

[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.



Figuur 7.1: Een schematische beschrijving van HFE (PW04).