On Asymptotic Security Estimates in XL and Gröbner Bases-Related Algebraic Cryptanalysis

Bo-Yin Yang, Jiun-Ming Chen, and Nicolas T. Courtois

[1] Tamkang University, Tamsui, Taiwan
by@moscito.org
[2] Chinese Data Security Inc., & Nat'l Taiwan U., Taipei
jmchen@math.ntu.edu.tw
[3] Axalto Smartcards, Paris, France
ncourtois@axalto.com

Abstract. "Algebraic Cryptanalysis" against a cryptosystem often comprises finding enough relations that are generally or probabilistically valid, then solving the resultant system. The security of many schemes (most important being AES) thus depends on the difficulty of solving multivariate polynomial equations. Generically, this is NP-hard.
The related methods of XL (eXtended Linearization), Gröbner Bases, and their variants (of which a large number has been proposed) form a unified approach to solving equations and thus affect our assessment and understanding of many cryptosystems.
Building on prior theory, we analyze these XL variants and derive asymptotic formulas giving better security estimates under XL-related algebraic attacks; through this examination we have hopefully improved our understanding of such variants. In particular, *guessing a portion of variables is a good idea for both XL and Gröbner Bases methods.*

Keywords: XL, Gröbner Bases, multivariate quadratics, algebraic cryptanalysis, asymptotic security estimates

1 Introduction

Modern cryptography relies critically on the difficulty to solve certain problems. RSA, currently dominant, depends on factoring; as integer factoring techniques improves and the speed of computers exponentiates ahead of embedded parts, it takes longer for smart cards to do modular arithmetic at a comfortable security level. Thus schemes relying on other hard problems are proposed. Solving generic multivariate polynomial systems is provably NP-hard ([22]), and many cryptosystems, including all multivariates (the input to the public maps are cut into small separate variables as opposed to treated as a big unit), depends on its difficulty. Systematic and algorithmic equations-solving have centered mostly around Gröbner Bases methods (cf. [2, 19, 20]). We will describe variants of the related XL method ([11]). Comparison shows FXL to be best, and that *guessing to make the equations suitably overdetermined is a generally good idea.*

Goal: To solve a (usually) quadratic system over a finite field $K = \mathrm{GF}(q)$. We denote the number of variables and equations by n and m respectively, and the equations are given as polynomials $\ell_1(\mathbf{x}) = \ell_2(\mathbf{x}) = \cdots = \ell_m(\mathbf{x}) = 0$.

J. López, S. Qing, and E. Okamoto (Eds.): ICICS 2004, LNCS 3269, pp. 401–413, 2004.

Procedure of Basic XL ([11]): Denote (per [42]) by $\mathbf{x}^{\mathbf{b}}$ the monomial $x_1^{b_1} x_2^{b^2} \cdots x_n^{b_n}$, and its total degree $|\mathbf{b}| = b_1 + \cdots + b_n$. $\mathcal{T} = \mathcal{T}^{(D)} = \{\mathbf{x}^{\mathbf{b}} : |\mathbf{b}| \leq D\}$ is the set of degree-D-or-lower monomials. Multiply each equation ℓ_i by all monomials $\mathbf{x}^{\mathbf{b}} \in \mathcal{T}^{(D-2)}$. Solving as a linear system $\mathcal{R} = \mathcal{R}^{(D)} = \{\mathbf{x}^{\mathbf{b}}\ell_j(\mathbf{x}) = 0 : 1 \leq j \leq m, |\mathbf{b}| \leq D-2\}$ in all the monomials $\mathbf{x}^{\mathbf{b}} \in \mathcal{T}^{(D)}$, or reduce the system to a univariate equation in some variable. The number of monomials will be denoted $T^{(D)} = T = |\mathcal{T}|$, total number of equations $R^{(D)} = R = |\mathcal{R}|$, and the number of independent equations $I^{(D)} = I = \dim(\mathrm{span}\mathcal{R})$.

Many claims have been made about XL and associated algebraic attacks. For applications to stream ciphers, see [7, 8]; to PKC's especially about its variant types, see [9, 12]; for block ciphers and the somewhat contentious related method of XSL, see [13], and also [29, 30]. We will build on recent theoretical developments in XL ([15, 36, 40, 42]) for better security estimates under XL and related methods, including Gröbner Bases. More details can also be found in [9, 12, 13].

1.1 A Framework for Estimating Security Levels

We need to solve a system or run a partial elimination. In this, we may apply any algorithm that takes time αN^ω to multiply two $N \times N$ matrices towards system-solving using Bernstein's Generalized Gaussian Elimination (GGE, [3]), with a time cost of (with β_0, β_1, β_2 depending on α and ω):

$$E_B(T, R) = \beta_0 R^\omega + \beta_1 T R^{\omega-1} + \beta_2 T^2 R^{\omega-2}.$$

While ω goes as low as 2.368 ([6]) in theory, practically we will use $\omega = \lg 7 \approx 2.8$, $\beta_0 = 1.16$, $\beta_1 = 48.5$, $\beta_2 = 0$ (cf. [3, 27, 35, 40]). If the system matrix can be blocked efficiently via coloring analysis ([16]), only the dominant block counts for T and R. When $m, n \to \infty$ so does R/T. However ([1, 5]) we can generate an almost-minimal equations randomly or via some extended-Buchberger type algorithm. So we will assume asymptotically $R/T \sim$ a constant.

On the other hand, then there is no way to beat sparse-matrix methods for finding a unique solution to a sparse system of equations. Standard estimates for Lanczos, Conjugate Gradients or Wiedemann methods ([17, 24, 38]) resemble

$$E_L(T, R) = (c_0 + c_1 \lg T)\, t\, T\, R,$$

where t counts the terms in each equation $(= \binom{n+2}{2})$, and c_0, c_1 are constants. We shall use an optimistic $c_0 = 16$, $c_1 = \frac{1}{4}$ for a complexity estimate in CPU cycles, which we should divide by 2^8 to get a rough estimate in 3DES blocks ([40] and simulations); the E_B estimate above is in contrast in field multiplications which is 2^{-6} of a 3DES block (the NESSIE unit, cf. [31]).

1.2 Basic Combinatorial Results Concerning XL

Over $K = \mathrm{GF}(q)$ Lemma 1 of [42] gives $T^{(D)} = [t^D]\ \left((1-t^q)^n (1-t)^{-(n+1)}\right)$, $R^{(D)} = mT^{(D-2)}$ (here $\underline{[u]s}$ is the $\underline{\text{coefficient of the monomial } u \text{ in the series}}$

expansion of s), so if D is roughly proportional to n, then so is $\lg T$. For large q (i.e. $q > D$), the above reduces to $T^{(D)} = \binom{n+D}{D}$ and for $q = 2$ to $T^{(D)} = \sum_{j=0}^{D} \binom{n}{j}$, so

Lemma 1. *If $D \sim wn$, then the Stirling formula and other asymptotics give*

$$\lg T \sim n\left[(1+w)\lg(1+w) - w\lg w\right] + o(n), \quad \textit{for large } q; \tag{1}$$

$$\sim n\left[-(1-w)\lg(1-w) - w\lg w\right] + o(n), \quad \textit{over } \mathrm{GF}(2); \tag{2}$$

$$\sim n\left[\lg \min(z^{-w}(1-z^q)/(1-z))\right] + o(n), \textit{ in general.} \tag{3}$$

This is why we mostly need only[1] D_0, the minimal operative degree D of XL:

Proposition 2 ([42]) *If equations ℓ_i are semi-regular, then for all $D < D_{reg}$,*

$$T - I = [t^D]\, G_{m,n}(t) = [t^D]\, \frac{(1-t^q)^n}{(1-t)^{n+1}} \left(\frac{1-t^k}{1-t^{kq}}\right)^m. \tag{4}$$

The degree of regularity $D_{reg} = \min\{D : [t^D]\, G_{m,n}(t) \leq 0\}$ is the smallest D such that Eq. 4 cannot hold if the system has a solution. The generating function $G_{m,n}(t)$ is also called the Hilbert Series. *$D_0 = \min\{D : [t^D]\, G_{m,n}(t) \leq 0\}$ is the (minimal) operative degree and usually equal to D_{reg}. Further, if the $(\ell_i)_{i=1\cdots m}$ are* non-semi-regular, *the value I can only decrease.*

Corollary 3 ([15, 42]). *$T - I = [t^D]\left((1-t)^{m-n-1}(1+t)^m\right)$ for generic quadratic equations if $D \leq \min(q, D^{\infty}_{reg})$, where D^{∞}_{reg} is the degree of the lowest term with a non-positive coefficient in the expansion of $G^{\infty}_{m,n}(t) = (1-t)^{m-n-1}(1+t)^m$. Again, the number I may only decrease for non-generic equations.*

Corollary 4. *With semi-regular quadratic equations over* GF(2)*, we have*

$$\textit{for all } D < D_{reg},\ T - I = [t^D]\, G^{(2)}_{m,n}(t) = [t^D]\left((1+t)^n(1+t^2)^{-m}(1-t)^{-1}\right).$$

So $D_0 = \min\{D : [t^D]\, G^{(2)}_{m,n}(t) \leq 1\} \approx D_{reg} = \min\{D : [t^D]\, G^{(2)}_{m,n}(t) \leq 0\}$.

Remark: If $T - I = [t^D]\, G_{m,n}(t)$ for all D, i.e., there are never any dependencies between the relations $\mathcal{R}$ other than those generated by $\ell_i[\ell_j] = \ell_j[\ell_i]$ and $\ell_i^{q-1}[\ell_i] = [\ell_i]$, then the polynomials ℓ_i form a *regular sequence*. This is obviously impossible if $m > n$, but the formula may yet hold until D is so large that the RHS of Eq. 4 becomes non-positive. This is the meaning of "no extraneous dependencies" in [28, 40], and semi-regularity in [2]: the XL-equations constructed according to an extension of the Buchberger criteria ([19]) have no extraneous interdependencies. Bardet *et al* (nor anyone else) give no general properties implying semi-regularity. As pointed out by C. Diem ([15]), the [42] proof is inaccurate. Commutative algebra has the concept of a sequence of polynomials being

[1] An anonymous reviewer for Crypto'04 opined that this is all the analysis necessary.

generic ([18]). It may be possible to prove Eq. 4 rigorously for generic polynomials ([15], using the *maximal rank conjecture* by R. Fröberg ([21]); Diem proves rigorously Cor. 3 in [15], and opines that Prop. 2 probably holds in general.

In any event, since it is also confirmed by many simulations ([1, 40, 42]) we will henceforth assume that Prop. 2 holds in general in the discussions below.

2 XL/FXL and Gröbner Bases with Guessing

We refine the estimates given for XL and FXL in [40] to use as a yardstick against which other XL variants can be measured. In the process we show that a suitable amount of guessing is generally useful with XL and Gröbner Bases.

2.1 The Old and the New: XL Estimates over Large Fields

It is known ([11, 15, 40]) if $f = m - n$ equal to (a) 0, then $D_0 = 2^n$ for $2^n < q$; (b) 1, then $D_0 = n + 1$; and (c) a constant ≥ 2, then $D_0 = n/2 - o(n)$. Indeed

Proposition 5 ([40]) *If fairly large m and q satisfy the premise to Cor. 3, then*

$$D_{reg} = \frac{m}{2} - (h_{f-1,1})\sqrt{\frac{m}{2}} + O(1) \sim \frac{m}{2} - \sqrt{fm}, \text{ for small } f(= o(\sqrt{m})). \quad (5)$$

$$= \left(\frac{1}{2} - \sqrt{c} + \frac{c}{2}\right) m + O(m^{\frac{1}{3}}), \text{ for } f = cm,\ m^{+\epsilon} \gg c \gg m^{-1/2-\epsilon}. \quad (6)$$

Here $f = m - n$ and $h_{k,1} = \sqrt{2k+1} + O(k^{-\frac{1}{6}})$ is the max. zero of the Hermite polynomial $H_k(x)$, known from analysis (cf. [37]). We also have

Proposition 6 *When $m = n < q < 2^m$, $D_0 \sim q + \frac{n}{2} - \sqrt{n}\,\mathrm{erfc}^{-1}(\frac{2}{n})$.*

Proof. If we assume generic quadratic equations, and that $D \geq q$ but is no larger than the smaller of $2q$ or D_{reg}, then Prop. 2 reduces to

$$T - I = [t^D]\ \left((1 - nt^q)(1-t)^{m-n-1}(1+t)^m\right).$$

Set $m = n$ and we may estimate $T - I$ by the Central Limit Theorem as

$$2^n - n\sum_{j=0}^{D-q}\binom{n}{j} \approx 2^n\left(1 - \frac{n}{\sqrt{2\pi}}\int_{-\infty}^{\frac{2(D-q-\frac{n}{2})}{\sqrt{n}}} e^{-\frac{u^2}{2}}du\right) = 2^m\left[1 - \frac{n}{2}\,\mathrm{erfc}\left(\frac{D-q-\frac{n}{2}}{\sqrt{n}}\right)\right],$$

where we use the complementary error function $\mathrm{erfc}x := \sqrt{\frac{2}{\pi}}\int_{-\infty}^{x} e^{\frac{-t^2}{2}}dt$.

Prop. 6 implies that when $m = n$ pure XL is infeasible for large dimensions. In the variant FXL, the attacker guesses at a number (denoted f) of variables and then runs XL, repeating until a solution is found. For XL or FXL with a given f, we can get very tight bounds ([40]) for D_{reg}. Since we have $\lg C_{\text{XFL}} \sim \omega \lg T + f \lg q$ where ω is the exponent of the elimination complexity (≈ 2.8 for Strassen-like methods), we may then use Eq. 6 to find (cf. also [40]) that

Proposition 7 *The optimal $c = f/n$ in FXL is the minimum point of*

$$(\lg q)c + \omega\left[(\tfrac{3}{2}-\sqrt{c}-\tfrac{c}{2})\lg(\tfrac{3}{2}-\sqrt{c}-\tfrac{c}{2}) - (\tfrac{1}{2}-\sqrt{c}+\tfrac{c}{2})\lg(\tfrac{1}{2}-\sqrt{c}+\tfrac{c}{2}) - (1-c)\lg(1-c)\right]$$

for sufficiently large q, and is denote $c_0 = c_0(q)$. This applies to $\mathbf{F_4}$-$\mathbf{F_5}$ *equally*[2].

Corollary 8. *Even when we start with $n/m = 1 - \epsilon + o(1)$ for a positive ϵ, the marginal cost of guessing is the same, so when $\epsilon \geq c_0$, then we should not guess any asymptotically significant portion of variables, but if $\epsilon < c_0$, we should guess at roughly $(c_0 - \epsilon)m = n - (1 - c_0)m$ more variables.*

For $q = 2^8$ and $\omega = 2$ (Lanczos) the minimum occurs at $c_0 = f/n \sim 0.049$, and we see that $\lg C_{\text{FXL}} \sim 2.4n$ (compared to $3.0n$ for $f = o(n)$). If $\omega = 2.8$ then the minimum is $\lg C_{\text{FXL}} \sim 3.0n$ (in contrast to $4.2n$ for $f = o(n)$) when $c_0 = f/n \sim 0.0959$. This proves that a suitable amount of guessing is a valuable concept in algebraic analysis of the XL-Gröbner-Bases family.

There is an alternative way to guess called XFL ([9, 42]) which delays guessing until the elimination has been performed on the highest-degree block of equations. What this does is effectively to lower D by 1, but its biggest drawback compared to FXL is *not* being compatible with Lanczos-like methods ([40, 42]).

Everyone ([12, 42] seems to consider FXL (and XFL) less than serious contenders for small fields, in particularly GF(2). But once we discard the notion that XL can be subexponential for roughly constant n/m, we will actually see, as below, that they are worthy all-around performers.

2.2 XL/FXL/XFL Estimates over Small Fields, Particularly GF(2)

To estimate the behavior of D_{reg} in small fields, one uses the method of Coalescent Saddle Point ([4, 23, 39]). Suppose we start with $m = n$, $q = 2$ (an equal number of quadratic equations and variables) and guess at $f = cn$ variables, then our asymptotic analysis starts with using Cauchy's Integration Formula:

$$[t^D]\, G^{(2)}_{n,n-f}(t) = [t^D]\left(\frac{1}{1-t}\frac{(1+t)^{n-f}}{(1+t^2)^{-n}}\right) = \frac{1}{2\pi i}\oint \frac{dz}{1-z}\left[\frac{(1+z)^{1-c}}{z^w(1+z^2)}\right]^n,$$

where $w := D/n$. At saddle points s the bracketed expression is stationary, so

$$\frac{1-c}{1+s} - \frac{2s}{1+s^2} - \frac{w}{s} = 0, \text{ or } (-c-1-w)\,s^3 + (-w-2)\,s^2 + (-c+1-w)\,s - w = 0. \quad (7)$$

Asymptotic behavior is determined at the saddle (stationary) points s. The method of Coalescent Saddle Points applies when we want an asymptotic expression to vanish, which means that the dominant term(s) of $g_{m,n}(D)$ must cancel, and this happens only when the cubic has double roots ([2, 4]):

$$4w^4 + (8+8c)\,w^3 + \left(8c^2 - 12c + 24\right)w^2 + \left(4c^3 - 16c^2 + 20\right)w + c^4 - 2c^3 - c^2 + 4c - 2 = 0,$$

[2] In the saddle-point computations of [2], one can easily seen that the coefficient of n in the asymptotic expansion of D_0 for $\mathbf{F_4}$-$\mathbf{F_5}$ is the same as that for XL/FXL, so the entire derivation carries over, as does most of this paper to $\mathbf{F_4}$-$\mathbf{F_5}$.

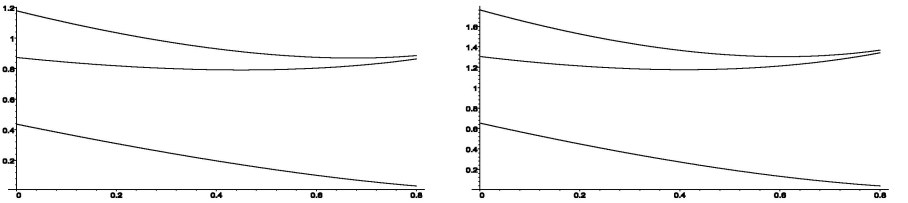
by taking the discriminant of the cubic. This is just like the behavior of $\mathbf{F_4}$-$\mathbf{F_5}$. We may write w via the Cardano-Ferrari formula or (when c is small) as a series:

$$w = D_{reg}/n = w_0(c) + O(n^{\frac{-2}{3}});\ w_0(c) = 0.0900 - 0.159\,c + 0.0568\,c^2 + 0.00800\,c^3 + O(c^4).$$

Given D_{reg}, we can estimate the complexity of the elimination phase via Eq. 2.

$$\lg C_{\text{XFL}/GGE} \sim (\omega\,[(1-c)\lg(1-c) - (1-c-w)\lg(1-c-w) - w\lg w] + c)\,n + o(n),$$
$$\lg C_{\text{FXL}/Lanczos} \sim (2\,[(1-c)\lg(1-c) - (1-c-w)\lg(1-c-w) - w\lg w] + c)\,n + o(n).$$

We plot $\lg T$ and w, the asymptotic coefficient of n in $\lg C$ against c, the proportion of variables fixed (looking at both Lanczos and GGE with the Strassen estimate) in Fig. 1(a). The minimum point is the optimal c, or $c_0 = c_0(q, \omega)$.



(a) Over GF(2) (b) Over GF(3)

Fig. 1. FXL/XFL cost (for $\omega = 2,\ \lg 7$) and w vs. Proportion of Variables Guessed

We can check that asymptotically, for Lanczos-like (resp. Strassen-like) methods we should fix (guess at) some $\sim 0.45n$ (resp. $0.67n$) variables for an asymptotic rate of $C \approx 2^{0.785n}$ (resp. $2^{0.865n}$) times a rational function. With "straight" XL or $\mathbf{F_5}$ we should get $C \approx 2^{0.873n}$ (resp. $2^{1.222n}$).

Example: We apply Prop. 2 and Sec. 1.1 instead of asymptotics. Suppose we have $m = n = 200$ over GF(2). Straight XL with Lanczos is expected to take about 2^{241} cycles while FXL with Lanczos, guessing at 120 (!) variables should be about 2^{199} cycles, already a little faster than brute-force. The GGE/Strassen case is even more exaggerated: straight XL should be 2^{305} multiplications; with 152 (!!) variables guessed FXL should be around 2^{206}, and XFL around 2^{196}.

For $q > 2$, we need more asymptotics, e.g., for $q = 3$, $\lg T/n$ is (up to $o(1)$):

$$\alpha \lg\left(\frac{\alpha\left(7\alpha - 3w + \sqrt{\alpha^2 + 6\alpha w - 3w^2}\right)}{2\,(2\alpha - w)^2}\right) - w\lg\left(\frac{-(\alpha - w) + \sqrt{\alpha^2 + 6\alpha w - 3w^2}}{2(2\alpha - w)}\right).$$

Hence we can get the similar plot for GF(3) in Fig. 1(b).

This means that asymptotically, XL methods (and its relatives the Gröbner bases methods) are faster than brute-force by rather more than was acknowledged in [2, 42], through the technique of guessing (FIXing). So FXL and XFL have their places for small fields too.

Comment: The above discussion is asymptotic; earlier disregard for FXL over small fields may well be justified for all practical dimensions.

3 Reassessing XL' and XLF

In [9], it was proposed that the variants XL' ([12]) and XLF can break SFLASH^{v2} ([34]) and the instance of HFE used in HFE Challenge 2. The former has $q = 2^7$, $m = n = 26$ (was 37, cut down with some guesses), and the latter $q = 2^4$, $m = n = 32$. It was pointed out ([40]) that the estimates in [9] were wrong.

We show below that, surprisingly, XL' and XLF are not likely to lead to asymptotically significant gains. In fact, they are asymptotically dominated by FXL. I.e., they lead to a higher leading-term (proportional to n) coefficient of the logarithm of the security complexity $\lg C$.

3.1 XL' and Its Security Estimates for Large q

XL' runs just like XL ([12]) except at the very end. Instead of coming down to one equation in one variable, we try to come down to a system of $\geq r$ equations in r variables and then solve by brute-force. We hope that this gives us a lower D than the D_0 for regular XL. In general we only require $T - I \leq \left([t^D](1-t)^{-r-1}(1-t^q)^r\right) - r$ with complexity

$$C_{\text{XL'}} \approx E\left(\binom{n+D}{D}, m\binom{n+D-2}{D-2}\right) + \frac{q^r D}{1-\frac{1}{q}}\binom{r+D}{D}. \tag{8}$$

Unfortunately, this decreased D may still not be small enough:

SFLASH^{v2}: [9] gave $D = 7$, $r = 5$. But here, $T - I = 3300336 \gg \binom{r+D}{D} = 792$ so XL' does not work. We actually need $D \geq 93$ ([1]).

HFE Challenge 2: [9] gave $D = 8$, $r = 10$, for which $T - I = 107594213 \gg \binom{r+D}{D} = 43758$. XL' may work at $D = 15$, $r = 19$ ([1]).

Indeed, it was established ([40]) that XL' is not very useful for large m, q and small $f = m - n$. In fact, when $f = 1$ or 2, XL' operates if and only if $D > m - r$; if $m = n$, XL' will not run at $D = m + 1 - r$, but will at $D = m + 2 - r$ for r large enough (around $r > m/2$). When r is small, we need a much larger D, around $2^{m/r}\,(r!)^{1/r}$. We show additionally below that XL' is asymptotically unsuitable for $f = m - n$ small, q large. What may be more surprising is that things do not get much better for small fields (Sec. 3.2).

Proposition 9 *For large n, q and any $f = m - n = o(n)$,*

1. *with $r/n = a + o(1)$, XL' needs at least the degree $D \sim n - r + o(n)$;*
2. *XL' does not lead to asymptotically significant gains.*

Proof. Assuming $D \sim wn$, $r \sim an$, asymptotically we can evaluate as follows:

$$T - I \sim \frac{\text{const}}{D^{f-1}}\binom{n+f}{D} \sim e^{n[-w\ln w-(1-w)\ln(1-w)+o(1)]}$$

$$\lesssim e^{n[(a+w)\ln(a+w)-w\ln w-a\ln a+o(1)]} \sim \binom{r+D}{D}.$$

or $(a+w)\ln(a+w) - a\ln a = -(1-w)\ln(1-w)$, which has the simple solution of $w = 1-a$. So the first part is proved. A reasonable approximation is that the minimum value of $C_{\text{XL'}}$ should happen when each of the elimination part and the brute-force part takes equal amount of time. I.e.

$$\omega\left((1+w)\ln(1+w) - w\ln w\right) = (-w\ln w - (1-w)\ln(1-w)) + (1-w)\ln q.$$

This has no solution between 0 and $\frac{1}{2}$ for most practical values of ω and q, so asymptotically for $f \geq 2$, XL' is not a significant improvement (because $D_{reg} \leq m/2$). Assuming $f = 1$ we can find numerical solutions: E.g., when $\omega = 2.8$, $q = 256$ we find that $w = 0.592$, which leads to $\lg C_{\text{XL'}} \sim 3.267m > \lg C_{\text{XFL}} \sim 3.00m$ over a fairly wide range of m (cf. Sec. 2.1).

3.2 Asymptotic Inefficiency of XL' for GF(2)

What is most surprising is that we can show that asymptotically, XL' does not work so great over GF(2) either, even though it was designed for that field. Let's assume that $m = n + o(n)$ and $D = wn$. We may compute the saddle points according to Eq. 7 (with $c = 0$). For any $w < 0.089979$ (the asymptotic limit for D_{reg}/n), Eq. 7 has three real roots of which we take the largest as s (we can verify this to be the dominant saddle point). The magnitude of $T - I$ can be approximated by $\lg(T-I) \sim n\left(\lg(1+s) - \lg(1+s^2) - w\lg s\right)$. If we run XL' with $r \sim am$ then

$$\frac{\lg(T-I)}{n} \sim \lg(1+s) - \lg(1+s^2) - w\lg s \approx a\lg a - w\lg w - (a-w)\lg(a-w) \sim \frac{\lg\binom{r}{D}}{n}$$

up to $o(1)$. If $w \approx 0.89979$ (i.e., $D \lesssim D_{reg}$), then $a \approx 0.85735$. Even for $\omega = \lg 7 \approx 2.8$, the brute-force searching stage would have a cost around $2^{1.273n}$, more than the XL/elimination cost of $2^{1.222n}$ (cf. Sec. 2.2). Suppose w decreases, a goes up and it gets even worse for XL'.

An obvious tweak is XFL': Combined XFL and XL'. Suppose we first fix $f \sim cn$ variables before doing XL' at $D \sim wn$ and $r \sim an$. However, as we repeat the computations above, we may verify that for c all the way up to 1, XFL' is still no improvement over XFL asymptotically.

Remark: Sec. 3.1–3.2 does *not* mean that XL'/XFL' is useless. It just says that XL' is unlikely to offer significant gain over GF(2) and by inference other small fields. There are still cases in which XL' will let us lower D by a little. If we are running XL/XFL with a Strassen-like elimination, we might as well pick up on XL' possibilities for free anyway.

3.3 XLF and Its Estimates

XLF ([9]) tries to utilize the Frobenius relations of $K = \text{GF}(q)$ when $q = 2^k$:

- Each generated XL equation in $\mathcal{R}$ is raised to the second, fourth,... powers easily (since this is a linear operation) as equations in $(x_i^2), \ldots, (x_i^{2^{k-1}})$, for k times as many variables *and* equations.
- Consider $(x_i^2), (x_i^4), \ldots, (x_i^{2^{k-1}})$ independent variables in addition to x_i. Use the fact that equivalent monomials are equal as new equations.

We know that ([40]) no more than $\Delta T = k\binom{n+\lfloor D/2\rfloor}{\lfloor D/2\rfloor} - 1$ extra equations are provided by XLF. Consequently when $D < q$, a necessary (and likely sufficient) operating condition for XLF is

$$[t^D]\left((1-t)^{m-n-1}(1+t)^m\right) - \binom{n+\lfloor D/2\rfloor}{\lfloor D/2\rfloor} < \lceil D/2\rceil. \tag{9}$$

Eq. 9 is how we can easily check that the cryptanalysis of [9] is nonfunctional.

The principal drawback ([1]) to XLF is that the dependencies are also copied k times. Indeed we can show this inefficiency to be intrinsic, i.e., when $f = m - n \leq 2$, we can prove ([40]) that XLF needs $D > n/2$ to operate. Of course this does not imply XLF to be useless, just less of an improvement than FXL, and that it does not lead to asymptotically significant gains.

Proposition 10 *XLF is asymptotically dominated by FXL.*

Proof. According to Eq. 9, for a fixed f and $D \sim wn$ we will have asymptotically

$$\begin{aligned}\lg(T-I) &\sim (-w\lg w - (1-w)\lg(1-w))\,n \\ &\gtrsim \left((1+\frac{w}{2})\lg(1+\frac{w}{2}) - \frac{w}{2}\lg(\frac{w}{2})\right)n \sim \binom{n+D/2}{n}.\end{aligned}$$

Assuming equality this yields as the only solution $w \approx 0.573$, which is greater than $w = 0.5$ from FXL/XFL. The complexity is $2^{4.17m}$ or $2^{2.98m}$ depending on whether Lanczos or Strassen is used, which is greater than that of FXL/XFL.

4 Further Discussions

Security levels of SFLASH and HFE challenge 2 under the variants of XL including FXL/XFL, XLF, and XL' can be updated using the formulas given in the text, and we tabulate them with some asymptotic estimates:

Table 1. XL estimates (3DES blocks): Previous ([9]) v. Bernstein ($\omega = 2.8$) v. Lanczos

XL Variant		XL	FXL	XL'	XFL	XLF
$n = 26$ $q = 2^7$ (SFLASH^{v2})	P	2^{282}	2^{82}	2^{58}	2^{71}	2^{67}
	B	2^{280}	2^{101}	2^{118}	2^{99}	2^{117}
	L	2^{208}	2^{85}	N/A	N/A	2^{92}
$n = 32$, $q = 2^4$ (HFE challenge 2)	P	2^{122}	N/A	2^{70}	2^{63}	2^{76}
	B	2^{151}	2^{97}	2^{115}	2^{93}	2^{145}
	L	2^{115}	2^{87}	N/A	N/A	2^{112}
Asymptotic for big n, q	L	$\binom{q+\frac{3n}{2}}{m}^{\omega}$	$2^{2.4n}$	N/A	N/A	$2^{3.0n}$
	B		$2^{3.0n}$	$2^{3.3n}$	$2^{3.0n}$	$2^{4.2n}$
Asymptotic in GF(2)	L	$2^{0.87n}$	$2^{.785n}$	N/A	N/A	N/A
	B	$2^{1.22n}$	$2^{.865n}$	$2^{1.27n}$	$2^{.865n}$	N/A

We will discuss a little about the remaining variant (one that we cannot quantify very well) before concluding.

4.1 A Brief Discourse on XL2 (and XSL)

This was first proposed ([12]) as an addendum to XL over GF(2), to add useful equations. The following formulation does not depend on $q = 2$, however. Let T' count the monomials that when multiplied by a given variable will still be in $\mathcal{T} = \mathcal{T}^{(D)}$. I.e. $T' = |\mathcal{T}_i'|$, where $\mathcal{T}_i' = \{\mathbf{x^b} : x_i\mathbf{x^b} \in \mathcal{T}\}$ for each i. Suppose I is not as large as $T - D$, but $C \equiv T' + I - T > 0$ (i.e. we have enough equations to eliminate all monomials not in $\mathcal{T}_i'$), then:

1. When doing elimination from the XL equations $\mathcal{R} = \mathcal{R}^{(D)}$, remove monomials not in $\mathcal{T}_1'$ first. We are then left with relations $\mathcal{R}_1$ that gives each monomial in $\mathcal{T} \setminus \mathcal{T}_1'$ as a linear combination of those monomials in $\mathcal{T}_1'$, plus C equations $\mathcal{R}_1'$ with only monomials in $\mathcal{T}_1'$.
2. Repeat for $\mathcal{T}_2'$ to get equations $\mathcal{R}_2$ and $\mathcal{R}_2'$ (we expect that $|\mathcal{R}_2'| = C$).
3. For each $\ell \in \mathcal{R}_1'$, use $\mathcal{R}_2$ to write every monomial in $\mathcal{T} \setminus \mathcal{T}_2'$ in the equation $x_1\ell = 0$ in terms of those in $\mathcal{T}_2'$. Do the converse for each $x_2\ell$, $\ell \in \mathcal{R}_2'$. We get $2C$ new equations.

Imai *et al* commented ([36]) that XL can be considered a variation of the $\mathbf{F}_4$-$\mathbf{F}_5$ algorithms and that XL2/XSL variants can be explained in terms of the Buchberger criteria ([36]). According to expert commentary ([1]), we should not restrict XL to two variables, but should operate on all variables at once, which resembles Gröbner Bases methods, and is generally consistent with the comment from [36]. We comment on these observations made in [42] about XL2 :

- Even when $I - (T - T') = C > 0$, XL need not run because some of the I independent equations may lack the monomials[3] in $\mathcal{T} \setminus \mathcal{T}_j'$.
- If $q > D$, XL2 operates when $T^{(D)} - I^{(D)} < T^{(D-1)} - (R^{(D-1)} - I^{(D-1)})$.

  We only need to eliminate the top monomials to run XL2 on all variables. That XL2 can run at least once is almost equivalent to XL on the homogeneous top-degree portion of the original equations terminating. For large q, this is essentially identical to having one fewer variable or rather, the requirement is that $G^{\infty}_{m,n-1}(D) = [t^D](1-t)^{m-n}(1+t)^m < 0$.

- Running XL2 at degree D on *all* variables x_i is equivalent to taking the relations $\mathcal{R}^{(D+1)}$ (i.e., the XL system created at degree $D+1$) and eliminate all the highest (degree-$(D+1)$) monomials to come down to degree D.
  Running XL2 to raise the degree by 1 does more work than FXL for large q.

  Going up one dimension results in between $2\times$ to $3\times$ as many equations and monomials with the dimensions we are working with. Doing XL2 with the entire n variables result in $n\times$ as many equations at the top level (substitution with known equations still takes the same order of time). In general it is only worthwhile if $n^{\omega-1}$ is less or equal to

[3] Example ([42]): Assume large q, $m = 11$, $n = 7$, and $D = 3$. We have $11\times(7+1) = 88$ equations in XL — all independent — and $\binom{7+2}{3} = 84$ cubic monomials, but *only* 77 *equations actually have cubic terms.*

$\left(\binom{n+d}{D+1}\binom{n+D-2}{D-1}^{\omega-1}\right) / \left(\binom{n+D-1}{d}\binom{n+D-3}{D-2}^{\omega-1}\right) = \frac{(n+D)(n+D-2)^{\omega-1}}{(d+1)(d-1)^{\omega-1}}$.

Note: We said nothing about going up 2 dimensions or more, but it start to resemble to an algorithm for Gröbner Bases; (cf. Sec. 4.2 below).

XSL is a modified XL construction when the equations are highly overdetermined, very sparse, and can be grouped cleanly into *S-Box'es* that share very few variables. Equation are multiplied only by monomials from other S-Boxes. Count the monomials and equations thus generated as T and I. XSL is harder to analyze and less well quantified. We point out a possible pitfall below, because XSL use thes "T' Method" as the last stage, which looks very much like XL2.

4.2 An Example Depicting the Pitfalls of Repeated XL2 Runs

Take any multivariate signature scheme with a 160-bit message or digest treated as 20 bytes, i.e., $q = 256$, $m = n = 20$. If memory is not a problem, using FXL with $f = 2$ and an optimistic estimate for the Lanczos algorithm, we expect to do 2^{72} 3DES blocks (2^{80} CPU operations, cf. [40]) and 100GB of RAM. The previous observations ([42]) means we cannot operate XL2 for $f = 2$. For $f = 3$, we can start operating XL2 at $D = 6$. We will find 13017 equations in $\mathcal{R}'$ that came from elimination between the degree-6 equations. If we multiply these by all variables, collate and eliminate again, we would have accomplished equivalent work to taking the XL system of equations (with $n = 17$, $m = 20$, $D = 7$) and eliminating all degree-7 and degree-6 monomials. Alas, there are insufficient equations for this purpose, hence XL2 cannot repeat in the same memory space.

So let us start with $D = 7$ instead. Using GGE with $\omega = 2.8$, the initialization takes about $\approx 2^{54}$ multiplications at the top block. We have 31350 equations that started at degree 6 or lower, plus 54093 equations that resulted from elimination on the degree-7 top block.

In running XL2, we must multiply a matrix of $54093 \times 17 = 919581$ rows and 245157 columns by a 245157×100947 to collate the equations, which takes about 2^{56} multiplications with suitable blocking, then eliminate down from a system of $919581 + 330600 = 1250181$ equations in 100947 variables (of which 85443 can be eliminated first at lower cost), that takes about 2^{57} multiplications or 2^{50} 3DES blocks. We *can* check that it is possible to eliminate down to degree $D = 6$ equations, for example because there are $I^{(8)} - I^{(6)} = 983535$ extra independent equations in going from degree 6 to 8, and only 980628 monomials of degree 7 and 8. The next XL2 step will take somewhat less than the above, and the total amount of time taken will be around 2^{58} multiplications per guess, or 2^{82} multiplications (2^{75} 3DES blocks) total. Other choices of f seem little better.

Comment: That XL2 can run once does not guarantee that it can repeat multiple times. This casts some doubt as to the applicability of the XSL attack.

4.3 Conclusion

Of the many XL variants, we have thus determined that FXL (XFL) is the best overall performer for very large n and relatively small $f = m - n$. Some of our

conclusions, such as those of Sec. 2.2, apply equally well to modernized Gröbner Bases Methods, because the two have very similar asymptotic characteristics. I.e., in $\mathbf{F_4}$-$\mathbf{F_5}$ over GF(2), we often really want to guess at a substantial proportion of the bit-variables before starting to run the algorithms. Guessing helps both memory and time requirements of the XL or Gröbner Bases algorithm.

Much remains still to be done in order to understand the impact of algebraic attacks on the security of very special systems such as derived from AES.

Acknowledgements

The first author would like to dedicate this work to the 60th birthday of his teacher and friend, Prof. Richard P. Stanley of MIT.

Note: Originally titled *Exact and Asymptotic Behavior of XL-Related Methods.*

References

1. Anonymous Referee Report from Crypto 2004.
2. M. Bardet, J.-C. Faugère, and B. Salvy, *Complexity of Gröbner Basis Computations for Regular Overdetermined Systems*, INRIA RR-5049.
3. D. Bernstein, *Matrix Inversion Made Difficult*, preprint at `http://cr.yp.to`.
4. C. Chester, B. Friedman, and F. Ursell, *An Extension of the Method of Steepest Descents,* Proc. Camb. Philo. Soc. 53 (1957) pp. 599–611.
5. D. Coppersmith, private communication.
6. D. Coppersmith, S. Winograd, *Matrix multiplication via Arithmetic Progressions*, J. Symbolic Computation, 9 (1990), pp. 251-280.
7. N. Courtois, *Higher-Order Correlation Attacks, XL Algorithm and Cryptanalysis of Toyocrypt*, ICISC '02, LNCS v. 2587, pp. 182–199.
8. N. Courtois, *Fast Algebraic Attacks on Stream Ciphers with Linear Feedback*, CRYPTO'03, LNCS v. 2729, pp. 177-194.
9. N. Courtois, *Algebraic Attacks over* GF(2^k)*, Cryptanalysis of HFE Challenge 2 and SFLASH^{v2}*, PKC '04, LNCS v. 2947, pp. 201-217.
10. N. Courtois, L. Goubin, and J. Patarin, *SFLASH^{v3}, a Fast Asymmetric Signature Scheme*, preprint available at `http://eprint.iacr.org/2003/211`.
11. N. Courtois, A. Klimov, J. Patarin, and A. Shamir, *Efficient Algorithms for Solving Overdefined Systems of Multivariate Polynomial Equations*, EUROCRYPT 2000, LNCS v. 1807, pp. 392–407.
12. N. Courtois and J. Patarin, *About the XL Algorithm over GF(2)*, CT-RSA 2003, LNCS v. 2612, pp. 141–157.
13. N. Courtois and J. Pieprzyk, *Cryptanalysis of Block Ciphers with Overdefined Systems of Equations*, ASIACRYPT 2002, LNCS v. 2501, pp. 267–287.
14. J. Daemen and V. Rijmen, *The Design of Rijndael, AES - The Advanced Encryption Standard.* Springer-Verlag, 2002.
15. C. Diem, *The XL-algorithm and a conjecture from commutative algebra*, ASIACRYPT 2004, to appear.
16. I. S. Duff, A. M. Erismann, and J. K. Reid, *Direct Methods for Sparse Matrices*, published by Oxford Science Publications, 1986.
17. W. Eberly and E. Kaltofen, *On Randomized Lanczos Algorithms*, Proc. ISSAC '97, pp. 176–183, ACM Press 1997.

18. D. Eisenbud, *Commutative Algebra with a View toward Algebraic Geometry*, Springer-Verlag 1995.
19. J.-C. Faugère, *A New Efficient Algorithm for Computing Gröbner Bases without Reduction to Zero (F5)*, Proceedings of ISSAC 2002, pp. 75-83, ACM Press 2002.
20. J.-C. Faugère and A. Joux, *Algebraic Cryptanalysis of Hidden Field Equations* (HFE) *Cryptosystems Using Gröbner Bases*, CRYPTO 2003, LNCS v. 2729, pp. 44-60.
21. R. Fröberg, An inequality for Hilbert Series of Graded Algebras, Math. Scand. 56(1985) 117-144.
22. M. Garey and D. Johnson, *Computers and Intractability, A Guide to the Theory of NP-completeness*, W. H. Freeman New York 1979.
23. Hsien-Kuei Hwang, *Asymptotic estimates of elementary probability distributions*, Studies in Applied Mathematics, 99:4 (1997), pp. 393-417.
24. B. LaMacchia and A. Odlyzko, *Solving Large Sparse Linear Systems over Finite Fields*, CRYPTO'90, LNCS v. 537, pp. 109–133.
25. D. Lazard, *Gröbner Bases, Gaussian Elimination and Resolution of Systems of Algebraic Equations*, EUROCAL '83, LNCS v. 162, pp. 146–156.
26. T. Matsumoto and H. Imai, *Public Quadratic Polynomial-Tuples for Efficient Signature-Verification and Message-Encryption*, EUROCRYPT'88, LNCS v. 330, pp. 419–453.
27. C. McGeoch, *"Veni, Divisi, Vici"*, Appearing in the "Computer Science Sampler" column of the Amer. Math. Monthly, May 1995.
28. T. Moh, *On The Method of XL and Its Inefficiency Against TTM*, available at `http://eprint.iacr.org/2001/047`
29. S. Murphy and M. Robshaw, *Essential Algebraic Structures Within the AES*, CRYPTO 2002, LNCS v. 2442, pp. 1–16.
30. S. Murphy and M. Robshaw, *Comments on the Security of the AES and the XSL Technique*, from author's homepage `http://www.isg.rhul.ac.uk/~sean/`
31. *NESSIE Security Report, V2.0*, available at `http://www.cryptonessie.org`
32. J. Patarin, *Hidden Field Equations* (HFE) *and Isomorphisms of Polynomials* (IP)*: Two New Families of Asymmetric Algorithms*, EUROCRYPT'96, LNCS v. 1070, pp. 33–48.
33. J. Patarin, L. Goubin, and N. Courtois, C^*_{-+} *and HM: Variations Around Two Schemes of T. Matsumoto and H. Imai*, ASIACRYPT'98, LNCS v. 1514, pp. 35–49.
34. J. Patarin, N. Courtois, and L. Goubin, *FLASH, a Fast Multivariate Signature Algorithm*, CT-RSA 2001, LNCS v. 2020, pp. 298–307. Update with SFLASH^{v2} available at `http://www.cryptonessie.org`
35. V. Strassen, *Gaussian Elimination is not Optimal*, Num. Math. 13(1969) pp. 354–356.
36. M. Sugita, M. Kawazoe, and H. Imai, *Relation between XL algorithm and Groebner Bases Algorithms*, preprint, `http://eprint.iacr.org/2004/112`.
37. G. Szegö, *Orthogonal Polynomials, 4th ed.*, publ.: Amer. Math. Soc., Providence.
38. D. Wiedemann, *Solving Sparse Linear Equations over Finite Fields*, IEEE Transaction on Information Theory, v. IT-32 (1976), no. 1, pp. 54–62.
39. R. Wong, *Asymptotic Approximations of Integrals*, Acad. Press (San Diego) 1989.
40. B.-Y. Yang and J.-M. Chen, *All in the XL Family: Theory and Practice*, preprint.
41. B.-Y. Yang and J.-M. Chen, *TTS: Rank Attacks in Tame-Like Multivariate PKCs*, available at `http://eprint.iacr.org/2004/061`.
42. B.-Y. Yang and J.-M. Chen, *Theoretical Analysis of XL over Small Fields*, ACISP 2004, LNCS v. 3108, pp. 277-288. *Note: updated version available from the authors.*