Open access • Book Chapter • DOI:10.1007/978-3-540-27800-9_24

Theoretical Analysis of XL over Small Fields — Source link

Bo-Yin Yang, Jiun-Ming Chen

Institutions: Tamkang University, National Taiwan University

Published on: 13 Jul 2004 - Australasian Conference on Information Security and Privacy

Topics: Serpent (cipher), Block cipher, Cryptanalysis, Advanced Encryption Standard and Multivariate cryptography

Theoretical Analysis of XL over Small Fields

Bo-Yin Yang[1] and Jiun-Ming Chen[2]

[1] Department of Mathematics, Tamkang University, Tamsui, Taiwan; by@moscito.org
[2] Chinese Data Security Inc. & National Taiwan U., Taipei, jmchen@math.ntu.edu.tw

Abstract. XL was first introduced to solve determined or overdetermined systems of equations over a finite field as an "algebraic attack" against multivariate cryptosystems. There has been a steady stream of announcements of cryptanalysis of primitives by such attacks, including stream ciphers (e.g. Toyocrypt), PKC's, and more controversially block ciphers (AES/Rijndael and Serpent).
Prior discussions of XL are usually heavy in simulations, which are of course valuable but we would like more attention to theory, because theory and simulations must validate each other, and there are some nuances not easily discerned from simulations. More effort was made in this direction of recent, but much of it was restricted to a large base field of size q, which is usually equal to 2^k.
By conducting an analysis of XL variants in general, we try to derive rigorous "termination conditions", minimal degree requirements for reliable, successful operation of XL and its relatives, hence better security estimates. Our work is applicable to small q, in particular the significant $q = 2$ case.
Armed with this analysis, we reexamine previously announced results. We conclude that XL and variants represent a theoretical advance that is especially significant over small fields (in particular over $\mathrm{GF}(2)$). However, its applicability and efficacy are occasionally overestimated slightly. We discuss possible future research directions. Much remains to be done.

Keywords: XL, finite field, multivariate cryptography, system of quadratic equations, algebraic attack.

1 Introducing the XL Family of Algorithms

XL is loosely descended from the relinearization ([17]) of Shamir and Kipnis. [8] implied that relinearization is superseded by XL which will always succeed if relinearization does. We will herein discuss only XL and its variants.

Goal: Find one solution to the system of m quadratic equations $\ell_1(\mathbf{x}) = \ell_2(\mathbf{x}) = \cdots = \ell_m(\mathbf{x}) = 0$ in n variables $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ over the base field $K = \mathrm{GF}(q)$. We will also use the following notations: The degree of a monomial $\mathbf{x}^{\mathbf{b}} = x_1^{b_1} x_2^{b_2} \cdots x_n^{b_n}$, is denoted $|\mathbf{b}| = \sum_i b_i$. The set $\mathcal{T} = \mathcal{T}^{(D)}$ comprises all monomials of total degree $\leq D$. It has $T = T^{(D)}$ elements.

1.1 Basic Procedures of XL

XL only operates on determined or over-determined systems, i.e. $n \leq m$. With more variables than equations, we must guess at enough variables so as to have at least as many equations as variables. XL at degree D then proceeds as follows:

1. "X" means to eXtend or multiply. Take all $\mathbf{x}^{\mathbf{b}} \in \mathcal{T}^{(D-2)}$ (i.e., all monomials of degree $\leq D-2$), and generate a set of equations $\mathbf{x}^{\mathbf{b}}\ell_i(\mathbf{x}) = 0$. The system of equations will be collectively termed $\mathcal{R} = \mathcal{R}^{(D)}$.
2. "L" means to Linearize. Run an elimination on the system of $R = mT^{(D-2)}$ equations $\mathcal{R} = \mathcal{R}^{(D)}$, treating each monomial $\mathbf{x}^{\mathbf{b}} \in \mathcal{T}^{(D)}$ as a variable. Enough equations must be independent to resolve the system. Because the system $\mathcal{R}$ is homogeneous in the variables (monomials) of $\mathcal{T}$, if there is a solution, then the number of independent equations (which we will denote I as opposed to $Free$ as in some earlier works) cannot exceed $T-1$, and is usually exactly $T-1$ if there is a unique solution. It was noted in [8] that I need not be as high as $T-1$. It suffices to be able to eliminate enough monomials to express 1 as a linear combination of powers of x_1 (or any other variable). Thus the *termination condition*, which ensures *reliable* resolution of the system, is $I = T - \min(D, q-1)$. This was first noted in [10].
3. *Solve the last remaining variable (usually x_1 as above), and recursively solve for the other variables as needed.* The time cost of XL is hence $C_{\text{XL}} = E(T, R)$, where $E(N, M)$ is *the cost elimination on N variables and M equations.*
   Courtois *et al* often uses $(7/64)\ T^{2.8}$ for $E(T, R)$ under Strassen's Blocking Elimination Algorithm when the field is $\text{GF}(2)$. We believe that an adjustment is needed. The best all-around elimination algorithm in the literature is [3], where the versatile D. J. Bernstein describes GGE, or the "Generalized Gaussian Elimination", a general way to compute what he termed a *quasi-inverse* to a non-square matrix. Via GGE we can solve an equation (his algorithm "S") or find a suitable basis of the kernel of a matrix, essentially a reduced echelon form (algorithm "N"). If the cost of multiplying two $N \times N$ matrices is $\sim \alpha N^{\omega}$, and $\gamma = 7\alpha/(2^{\omega}-4)$ then the time cost of GGE is given by

$$E_S(N, M) = \frac{2\alpha(1+\gamma)}{(2^{\omega}-2)} M^{\omega-1} N + \frac{\alpha M^{\omega}}{(2^{\omega}-1)}; \tag{1}$$

$$E_N(N, M) = \frac{2\alpha(2+\gamma)}{(2^{\omega}-2)} M^{\omega-2} N^2 + \frac{4\alpha\gamma}{7} M^{\omega-1} N + \frac{\alpha M^{\omega}}{(2^{\omega}-1)}. \tag{2}$$

   With $\alpha = 7/64$ we get $E_S(T, R) = 0.76TR^{1.8} + 0.023R^{2.8}$. This likely represents a better estimate than $(7/64)T^{2.8}$, *because Strassen's original algorithm has a large probability of failure particularly in* $\text{GF}(2)$*, and even the later Bunch-Hopcroft ([2]) version can work only for square matrices.*

Note: We need to decide on D before the algorithm is run, hence this study.

[8] assumed D_0, *the minimum D needed for XL to work*, to be not far removed from what makes $R > T$, and hence obtained the heuristic of $D_0 \sim n/\sqrt{m}$, and [10] repeated this estimate for $\text{GF}(2)$. Over $\text{GF}(2)$ XL will work with most dimensions, but for a large q it was found that $D_0 = 2^n$ for $m = n$; and[3] $D_0 = n = m-1$ when $n = m-1$. [8] claimed that when $n \leq m-2$, "it is likely" that $D_0 \approx \sqrt{n}$ because D_0 "drops abruptly when $m-n$ increases", although [7] verified that for larger dimensions, *"$m-n$ may need to be yet higher"*. So XL sometimes work less smoothly, which led to FXL ([8]). We seek to understand the behavior of D_0 (hence XL and variants) better.

[3] The formula below may be off by one equation, i.e. the authors may have meant m.

1.2 The XL' Variant

XL' operates like XL ([10]), except that we try to eliminate down to r equations that involves only monomials in r of the variables, say $x_1, \ldots, x_r$, then solve the remaining system by brute-force substitution. If T' is the number of degree $\leq D$ monomials in r variables, then we require $T - I$ to be at most $T' - r$ instead of D. Hopefully we can run with a smaller D, and q^r is relatively small. The time complexity is then bounded by

$$C_{\text{XL'}} = E(T, R) + q^r T' D / \left(1 - \tfrac{1}{q}\right). \tag{3}$$

Note that we must test degree-D polynomials with r variables and up to $\binom{r+D}{D}$ terms, and there is a $1/q$ probability for any polynomial to vanish on random inputs. *We will discuss the behavior of XL' in the next section more accurately.*

1.3 The XL2 Variant: New Equations from Old?

Let $\mathcal{T}_i'$ be the set of monomials that when multiplied by x_i will still be in $\mathcal{T} = \mathcal{T}^{(D)}$, and T' be their number. I.e. $T' = |\mathcal{T}_i'|$, where $\mathcal{T}_i' = \{\mathbf{x}^{\mathbf{b}} : x_i \mathbf{x}^{\mathbf{b}} \in \mathcal{T}\}$. If $T > I$, $C \equiv T' + I - T > 0$, then we can try to generate more useful equations:

1. Starting from the equations $\mathcal{R} = \mathcal{R}^{(D)}$, we eliminate monomials not in $\mathcal{T}_1'$ first. We are then left with relations $\mathcal{R}_1$, which gives each monomial in $\mathcal{T} \setminus \mathcal{T}_1'$ as a linear combination of monomials in $\mathcal{T}_1$, plus C equations $\mathcal{R}_1'$ with terms only in $\mathcal{T}_1'$.
2. Repeat for $\mathcal{T}_2$ to get the equations $\mathcal{R}_2$ and $\mathcal{R}_2'$ (we should also have $|\mathcal{R}_2'| = C$).
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. [10], which proposed XL2 over $\mathrm{GF}(2)$ only, suggests that most of these $2C$ equations will be linearly independent, because they are somehow built out of all the equations. It was also remarked that XL2 can be repeated as needed for more equations and eventually a solution. The attacker can also run XL2 using more variables, as in adding $(x_3 \mathcal{R}_3')$, and so on.

1.4 Other Relatives of XL: FXL, XFL, XLF, and XSL

FXL and XFL: The "F" here means to "fix" ([8]). f variables are guessed at random in the hope that the degree D_0 needed for XL will decrease. After guessing at each variable, XL is run and tested for a valid solution. In [7], it was proposed that the R equations be generated and an elimination be run on them as far as it can go *before* guessing at the variables. It is named *improved FXL*, but we think that *XFL* suits better. For large (m, n) these are mainly useful (cf. Sec. 3) for removing excess solutions. *To cut the operative D, we need f proportional to n (cf. [24]).*

XLF: In [7] the authors proposed a variation, trying to use the Frobenius relations $x^q = x$ to advantage when $q = 2^k$, by considering (x_i^2), (x_i^4), …, $(x_i^{2^{k-1}})$ as new independent variables, replicating all R equations k times by repeatly squaring them, and using the equivalence of identical monomials as extra equations.
The variant is called XLF, for "field" or "Frobenius equations".

XSL: Not a true XL relative, this is a related linearization-based method designed to work on *overdefined systems of sparse quadratic equations that characterize certain block ciphers*. [9] suggested that it may be possible to break AES using XSL and thereby raised a storm of controversy. Occasionally we see amazingly low numbers given for this attack based on applying XSL to structural equations discovered by Murphy and Robshaw ([19]) in AES, but in contrast to the general public, few researchers appear to believe that AES has been broken. We mention XSL only because its final stage or the 'T'-method" resembles XL2.

It was thought that FXL/XFL/XLF for small fields like $\mathrm{GF}(2)$ *do not appreciably increase speed compared to original XL. But we refer you to [24] for an update.*

2 Termination Behavior for XL – a Combinatorial Study

We discuss when XL can be expected to terminate using combinatorial technique. We first prove an easy lemma about T in general. The combinatorial notation $\underline{[u]\,p}$ will denote '$\underline{\text{the coefficient of term } u \text{ in the expansion of } p}$". E.g. $[x^2](1+x)^4 = 6$.

Lemma 1 (Number of Monomials up to a Given Degree).

$$T = T^{(D)} = [t^D]\frac{(1+t+t^2+\cdots+t^{q-1})^n}{1-t} = [t^D]\frac{(1-t^q)^n}{(1-t)^{n+1}} \tag{4}$$

Proof. Consider the product $\prod_{i=1}^{n}(1+x_i+x_i^2+\cdots x_i^{q-1}) = \prod_{i=1}^{n}[(1-x_i^q)/(1-x_i)]$. This generates all possible monomials, exactly once each. Set every x_i be equal to t in this expression, and clearly the coefficient of t^D counts the monomials of degree exactly D. As $(a_0 + a_1 t + a_2 t^2 + \cdots + a_D t^D + \cdots)(1+t+t^2+\cdots+t^D+\cdots)$ has as its D-th degree coefficient $(a_0 + \cdots + a_D)$, we have derived $T^{(D)}$ henceforth. □

Lemma 1 unites as useful corollaries the special cases of large D (i.e. $D > q$) where $T = \binom{n+D}{D}$ as first given in [8], and $q = 2$ when $T = \sum_{i=0}^{D}\binom{n}{i}$ as in [10]. $R^{(D)} = mT^{(D-2)}$ and hence is given by $m[t^{D-2}]\left((1-t^q)^n/(1-t)^{n+1}\right)$.

We want to know how many independent equations there are in the general case, when dependencies abound among the equations. Denote by $[f]$ the equation $f(\mathbf{x}) = 0$, and assume that $\ell_i(\mathbf{x}) = \sum_{j\le k} a_{ijk}x_jx_k + \sum_j b_{ij}x_j + c_i$, then

$$\sum_{j\le k} a_{ijk}[x_jx_k\ell_{i'}] + \sum_j b_{ij}[x_j\ell_{i'}] + c_i\,[\ell_{i'}] = \sum_{j\le k} a_{i'jk}[x_jx_k\ell_i] + \sum_j b_{i'j}[x_j\ell_i] + c_{i'}[\ell_i],$$

I.e. $[\ell_i\ell_j]$ appears as two different linear combinations of the equations. And there will be dependencies among the dependencies, so it is not so obvious that we can compute the number of free equations under reasonable conditions.

Remark: T. Moh's critique on XL ([18]) did not clarify what conditions or 'extraneous dependencies" may be, stating only 'from well-known facts in algebraic geometry". Mathematically, what Moh implied adds up to the ℓ_i's forming a *semi-regular sequence* (see [1] for a complete definition compatible with Moh's results). [11] is likely the most

rigorous of the independent derivations to date. The validity of many generating series result on XL and $\mathbf{F_4}$-$\mathbf{F_5}$ rests on the *Maximum Rank Conjecture* ([15]). This implies that (among other things, cf. [11]) for a generic sequence (ℓ_i) in a infi nite fi eld, the sum of the ideals $\{f(\mathbf{x}) : (\ell_i\ell_j)|f\}$ for $i < j$ is equal to $\{f(\mathbf{x}) = \sum_i p_i(\mathbf{x})\ell_i(\mathbf{x}) : \ell_j|f\}$. We will assume that this holds in our slightly different case.

Theorem 2. *The number of independent XL equations over* $\mathrm{GF}(q)$ *is bound by*

$$T - I \geq [t^D] \left(\frac{1}{1-t}\left(\frac{1-t^q}{1-t}\right)^n \left(\frac{1-t^2}{1-t^{2q}}\right)^m\right), \text{ for all } D < D_{reg}, \qquad (5)$$

where D_{reg} *is the "degree of regularity" defined by*

$$D_{reg} = \min\{D : [t^D] \left((1-t)^{-n-1}\,(1-t^q)^n\,(1-t^2)^m\,(1-t^{2q})^{-m}\right) \leq 0\}. \qquad (6)$$

If there are no extra dependencies, the bound would be an equality. For this to happen, no $\ell_i(\mathbf{x}) - \alpha$ *can be non-trivially factorizable for any* ℓ_i *and* $\alpha \in \mathrm{GF}(q)$*, and the* ℓ_i*'s must contain enough degree-2 monomials so make* $\deg \ell_i\mathbf{x}^{\mathbf{b}} = |\mathbf{b}| + 2$ *for any* $\mathbf{b}$ *and any* i*. These conditions being met, the minimum* D *for XL to operate reliably is*

$$D_0 = \min\left\{D : [t^D]\,\frac{1}{1-t}\left(\frac{1-t^q}{1-t}\right)^n \left(\frac{1-t^2}{1-t^{2q}}\right)^m \leq \min(D, q-1)\right\}. \qquad (7)$$

Proof. Linear subspaces of $\mathrm{span}\mathcal{T}^{(\mathcal{D})}$ (a.k.a. degree $\leq D$ polynomials in $x_1, \ldots, x_n$) form a partially ordered set. In fact, the intersection and the algebraic sum ($U + V \equiv \{\mathbf{u}+\mathbf{v} : \mathbf{u} \in U,\ \mathbf{v} \in V\}$) fulfi ll requirements for the infi mum and supremum operations of a *modular lattice* with the dimension as the rank function, which in plainer language means that for any subspaces U and V, we have

$$\dim U + \dim V = \dim(U + V) + \dim(U \cap V). \qquad (8)$$

Eq. 8 implies (cf. previous section and [11]) a form of the Principle of Inclusion-Exclusion ([21]), one which states that if (a) A_i is the subspace of $\mathrm{span}\mathcal{T}^{(D)}$ comprising all polynomials of degree $\leq D$ divisible by ℓ_i; and (b) the ℓ_i form a semi-regular sequence, then

$$\dim\left(\sum_{i=1}^{k} A_i\right) = \sum_{j=1}^{k}(-1)^{j-1}\left(\sum_{1\leq i_1<\cdots<i_j\leq k} \dim(A_{i_1} \cap \cdots \cap A_{i_j})\right). \qquad (9)$$

Let A_i be the set comprising all polynomials of degree $\leq D$ that is divisible by ℓ_i. So $\mathrm{span}\mathcal{R}^{(D)} = \sum_{i=1}^{m} A_i$ and $I = \dim \mathrm{span}\mathcal{R}^{(D)}$, and we need to compute $\dim A_i$, $\dim A_i \cap A_j$, and in general $\dim A_{i_1} \cap \cdots \cap A_{i_j}$. We fi rst fi nd the dimension of $A_i = \mathrm{span}\,\{\mathbf{x}^{\mathbf{b}}\ell_i : |\mathbf{b}| \leq D-2\} = \ell_i\,\mathrm{span}\,(\mathcal{T}^{(D-2)})$. We would have $\dim A_i = T^{(D-2)}$ *were there not* $f(\mathbf{x})[\ell_i(\mathbf{x})]$ *of degree* $\leq D$ *that are identically zero.* Since ℓ_i is assumed not to factor, $f(\mathbf{x})[\ell_i(\mathbf{x})]$ vanishes *iff* $f(\mathbf{x})$ is divisible by $(\ell_i(\mathbf{x}))^{q-1} - 1$, hence we have a 1-to-1 correspondence of A_i with a quotient space of $\mathrm{span}\,\mathcal{T}^{(D-2)}$ by $\{f(\mathbf{x}) : \deg f \leq D-2,\ (\ell_i^{q-1}-1)|f\} = (\ell_i^{q-1}-1)\,\mathrm{span}\,\mathcal{T}^{(D-2q)}$. Ergo,

$$\dim A_i = \dim \ell_i\,\mathrm{span}\,(\mathcal{T}^{(D-2)}) = T^{(D-2)} - \dim\left((\ell_i^{q-1}-1)\,\mathrm{span}\,(\mathcal{T}^{(D-2q)})\right).$$

So $\dim A_i$ would be $T^{(D-2)} - T^{(D-2q)}$ *except for that* $(\ell_i^{q-1}-1)\,\mathrm{span}\,(\mathcal{T}^{(D-2q)}) = \{(\ell_i^{q-1}-1)g(\mathbf{x}) : \deg g \le D-2q\}$, is not in bijective correspondence with $\mathrm{span}\,(\mathcal{T}^{(D-2q)})$ because we have to discount (or quotient out) all g such that $(\ell_i^{q-1}-1)g(\mathbf{x}) = 0$. Under the conditions the theorem, this means $\ell_i | g$, so we must recompensate to get

$$\dim A_i = T^{(D-2)} - T^{(D-2q)} + \dim\left(\ell_i\,\mathrm{span}\,(\mathcal{T}^{(D-2q-2)})\right).$$

By the same reasoning, we must deduct $T^{(D-4q)}$, add $T^{(D-4q-2)}$, and so on, repeating until we hit zero. If we let $p(t) = (1-t)^{-n-1}(1-t^q)^n$, then $T^{(D)} = [t^D]p(t)$ and

$$\begin{aligned}\dim A_i &= [t^{D-2}]p - [t^{D-2q}]p + [t^{D-2q-2}]p - [t^{D-4q}]p + [t^{D-4q-2}]p - + - + \cdots \\ &= [t^D]\left((t^2 - t^{2q})\,p/(1-t^{2q})\right). \qquad (10)\end{aligned}$$

$$\text{Similarly,}\quad \dim\left((\ell_i^{q-1}-1)\mathrm{span}\,\mathcal{T}^{(D-2q)}\right) = [t^D]\left(t^{2q}(1-t^2)\,p/(1-t^{2q})\right). \qquad (11)$$

What is $\dim A_i \cap A_j = \dim\left(\ell_i\ell_j\,\mathrm{span}\,\mathcal{T}^{(D-4)}\right)$? $\ell_i(\ell_i^{q-1}-1) = \ell_j(\ell_j^{q-1}-1) = 0$, so any $(\ell_i^{q-1}-1)g_1(\mathbf{x}) + (\ell_j^{q-1}-1)g_2(\mathbf{x})$ vanishes when multiplied by $\ell_i\ell_j$. Hence

$$A_i \cap A_j \cong \mathrm{span}\,\mathcal{T}^{(D-4)} \Big/ \left((\ell_i^{q-1}-1)\,\mathrm{span}\,\mathcal{T}^{(D-2q-2)} + (\ell_j^{q-1}-1)\,\mathrm{span}\,\mathcal{T}^{(D-2q-2)}\right).$$

So $\dim A_i \cap A_j$ should be $T^{(D-4)}$ minus the dimension of the subspace spanned by multiples of $(\ell_i(\mathbf{x})^{q-1}-1)$ and $(\ell_j(\mathbf{x})^{q-1}-1)$ that are of degree $\le D-4$, and

$$\begin{aligned}\dim A_i \cap A_j = T^{(D-4)} &- \dim\left((\ell_i^{q-1}-1)\,\mathrm{span}\,\mathcal{T}^{(D-2q-2)}\right) \\ &- \dim\left((\ell_j^{q-1}-1)\,\mathrm{span}\,\mathcal{T}^{(D-2q-2)}\right) + \dim\left((\ell_i^{q-1}-1)\,(\ell_j^{q-1}-1)\,\mathrm{span}\,\mathcal{T}^{(D-4q)}\right),\end{aligned}$$

again via Eq. 8, and $\dim\left((\ell_i^{q-1}-1)\,(\ell_j^{q-1}-1)\,\mathrm{span}\,\mathcal{T}^{(D-4q)}\right)$ via the same reasoning is found to be $T^{(D-4q)}$ minus

$$\dim\left(\ell_i\,\mathrm{span}\,\mathcal{T}^{(D-4q-2)}\right) + \dim\left(\ell_j\,\mathrm{span}\,\mathcal{T}^{(D-4q-2)}\right) - \dim\left(\ell_i\ell_j\,\mathrm{span}\,\mathcal{T}^{(D-4q-4)}\right).$$

I.e., if $\dim A_i \cap A_j = [t^D]f(t)$, then it is also equal to

$$[t^D]\left((t^4p) - 2\left(\frac{t^{2q+2}(1-t^2)}{1-t^{2q}}p\right) + (t^{4q}p) - 2\left(\frac{t^{4q}(t^2-t^{2q})}{1-t^{2q}}p\right) + \left(t^{4q}f(t)\right)\right),$$

which routinely simplifies to $[t^D]f(t) = \dim A_i \cap A_j = [t^D]\left((t^2-t^{2q})^2p/(1-t^{2q})^2\right)$. The same Inclusion-Exclusion manuever assisted by mathematical induction shows that

$$\dim\left(A_{i_1} \cap \cdots \cap A_{i_j}\right) = [t^D]\left(\left(\frac{t^2-t^{2q}}{1-t^{2q}}\right)^j p\right); \quad \text{in fact, we have also} \qquad (12)$$

$$\dim\left(\prod_{k=1}^{j}(\ell_{i_k}^{q-1}-1)\cdot\,\mathrm{span}\,\mathcal{T}^{(D-2jq)}\right) = [t^D]\left(\left(\frac{t^{2q}-t^{2q+2}}{1-t^{2q}}\right)^j p\right). \qquad (13)$$

To see that this is so, we apply Eq. 8 two more times in succession on

$$\dim\left(A_{i_1}\cap\cdots\cap A_{i_j}\right)=T^{(D-2j))}-\dim\left(\sum_{k=1}^{j}\left(\ell_{i_k}^{q-1}-1\right)\operatorname{span}\mathcal{T}^{(D-2(q+j-1))}\right).$$

and we verify Eqs. 12 and 13 as consistent. Substituting Eq. 12, we finally get

$$\begin{aligned}T-I&=[t^D]\left(p\sum_{k=0}^{m}(-1)^k\binom{m}{k}\,(t^2-t^{2q})^k/(1-t^{2q})^k\right)=[t^D]\left(p\,\left(1-\frac{t^2-t^{2q}}{1-t^{2q}}\right)^m\right)\\&=[t^D]\,\left((1-t)^{-n-1}(1-t^q)^n\,(1-t^2)^m\,(1-t^{2q})^{-m}\right).\quad\text{[=Eq. 5]}\end{aligned}$$

This cannot hold if the right hand side is non-positive, which also indicates that XL will terminate. With a conjecture in commutative algebra, we can show that I will only be smaller ([11]). □

Corollary 3. *When applying XL over the field* $\mathrm{GF}(2)$ *at degree* $D<D_{reg}$*, then*

$$T-I=[t^D]\,\left((1-t)^{-1}\,(1+t)^n\,(1+t^2)^{-m}\right). \tag{14}$$

XL will usually terminate if the RHS≤ 0*, with a unique solution if* $T-I=1$.

Note: Eq. 14 is consistent with the partial results (i.e. for $D\leq 5$) of [10].

Corollary 4. *The degree of regularity (the maximum degree of in the elimination stage) of the Gröbner algorithm* $\mathbf{F_5/2}$ *is no lower than XL's degree of regularity over* $\mathrm{GF}(2)$.

Proof. The degree of regularity for $\mathbf{F_5/2}$ ([13, 14]) is given (according to [1]):

$$D_{reg}^{F_5/2}=\min\left\{D:[t^D]\left(\frac{(1+t)^n}{(1+t^2)^m}\right)\leq 0\right\}. \tag{15}$$

$D_{reg}^{F_5/2}$ is the degree of the first non-positive coefficient in $(1+t)^n(1+t^2)^{-m}$. Compare this to Eq. 6, where our $D_{reg}^{\mathrm{XL}/2}$ is the degree of the first non-positive coefficient in $(1+t)^n(1+t^2)^{-m}(1-t)^{-1}$, or the lowest degree up to which the coefficients sum up non-positive in $(1+t)^n(1+t^2)^{-m}$. We see that $D_{reg}^{F_5/2}\leq D_{reg}^{\mathrm{XL}/2}$. Actually,

$$D_{reg}^{\mathrm{XL}/2}\gtrsim D_{reg}^{F_5/2}\sim 0.900m+O(m^{\frac{1}{3}}), \tag{16}$$

via the same kind of asymptotic argument in [1]. *Note that the operation of both* $\mathbf{F_5/2}$ *and* XL*/2 are hinged on the sparse system solving stage, so we cannot conclude that* $\mathbf{F_5/2}$ *is faster even though it will have a smaller* D_0. □

Corollary 5. *XL' (cf. Sec. 1.2) applied over* $\mathrm{GF}(2)$ *will operate when the* $T-I\leq\sum_{i=0}^{D}\binom{r}{D}-r$ *(cf. Eq. 5). This reduces to Theorem 2 if* $r=1$*, as expected.*

The large q case differs sufficiently from small q that all further XL discussions on large fields probably belongs to another paper ([23]). But merely setting $t^q=0$ we get:

Corollary 6 (Large q case). *For large q and $D < D_{reg}$, we have*

1. *If $q > D$, then $T - I = [t^D]\ \left((1-t)^{m-n-1}\ (1-t^2)^m\right) \geq 0$.*
2. *D_{reg} (resp. D_0) is the least D such that the RHS above ≤ 0 (resp. $\leq \min(D, q-1)$).*
3. *If $\max(2q, D_0) > D \geq q$, then $T - I = [t^D]\left((1-t)^{m-n-1}\ (1-nt^q)\ (1-t^2)^m\right)$.*

Theorem 2 carries over nicely to higher-order equations with little change.

Theorem 7 (Non-Quadratic Equations). *If $\deg \ell_i = d_i$ instead of 2, then*

$$T - I = [t^D]\ \left(\left(\frac{(1-t^q)^n}{(1-t)^{n+1}}\right)\ \prod_{i=1}^{m}\left(\frac{(1-t^{d_i})}{(1-t^{d_i q})}\right)\right), \quad \textit{for all } D < D_{reg}. \tag{17}$$

assuming no extra dependencies. Just as in Theorem 2, D_{reg} is the smallest D for which the left side of Eq. 17 is non-positive. In particular, if $\deg \ell_i = k$ for all i, then

$$T - I = [t^D]\ \left(\frac{1}{1-t}\ \left(\frac{1-t^q}{1-t}\right)^n\ \left(\frac{1-t^k}{1-t^{kq}}\right)^m\right), \quad \textit{for all } D < D_{reg}. \tag{18}$$

The proof carries so well, in fact, that we can see that *if for example one of the ℓ_i is a product of two factors of degree k and k', then the corresponding factor in Eq. 17 becomes* $\left(1 - \left((t^k - t^{kq})(t^{k'} - t^{k'q})\right) / \left((1-t^{kq})(1-t^{k'q})\right)\right)$.

This theorem governs the behavior of XL when used for generalized or higher-order correlation attacks such as in [5, 6], which is an application of Eq. 18 with $k = 3$.

3 Looking at Earlier Claims and Results over GF(2)

Enough theory! We turn to some practical assessment of XL over GF(2). An interesting tidbit from Eqs. 14 and 16 is that when $m = n$ (in fact whenever $m/n \to \beta$, a constant), we do not have $D_0 \approx n/\sqrt{m}$ as postulated by [10]; instead, Eq. 16 gives $D_0 \sim cn$, where $c \approx 0.09$ for $m = n$. Let us back this up by plotting up to around 2000:

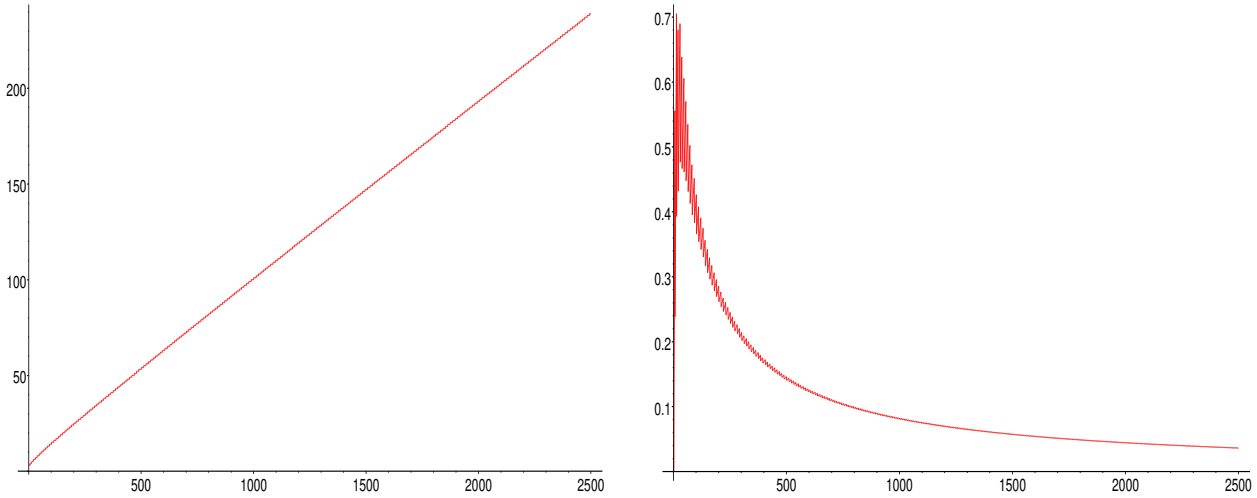
The D_0 vs. m graph is a straight line with a linear correlation coeffi cient of around 0.9999. This precludes the ratio $\mu = T/R$ at $D = D_0$ from being proximate to 1. It apparently decreases to 0 inversely to an increasing m, which will be proved in Prop. 8.

[4] claims to break HFE ([20]) challenge 1 ($n = m = 80$). This is a "generalized algebraic attack", not via XL over GF(2). In [10] the authors were more cautious, allowing that for $m = n$ XL may turn out to be always somewhat slower than brute-force search. Using Eq. 6, $D_0 = 12$ for $m = n = 80$. The number of operations (by the Bernstein formula) is $\approx 2^{177}$ ($\approx 2^{154}$ substitutions), well above brute-force search.

All is not completely lost for XL. Solving sparse equations (cf. related texts, e.g. [12]) is easier than in general. If there are N variables, M rows, but only t terms in each row, an optimistic bound for a time cost is $E_L = tMN(c_0 + c_1 \lg N)$. The constants c_0 and c_1 for the best case are usually around 10 and $1/10$ respectively (and can be several times more). This get us a complexity of around 2^{84} substitutions, which is much better.

Let us justify somewhat the assessment of [5] of exponential running time for XL.

Proposition 8. *For $m = n \to \infty$, XL runs in exponential time ($\sim 2^{7n/8}$) over GF(2).*



(a) Depicting D_0 with respect to $m = n$

(b) $\mu = T/R$ at D_0 w.r.t $m = n$

Fig. 1. Behavior of XL over $\mathrm{GF}(2)$ when $m = n$

Proof. If $m \to \infty$ while $\frac{D}{m} \to \alpha \approx 0.09$, then $\frac{R}{T} = \frac{m\binom{m}{D-2}}{\binom{m}{D}} \to \frac{m}{(\frac{1}{\alpha}-1)^2} \approx \frac{m}{100}$, and

$$T \sim \sum_{j=0}^{\alpha n} \binom{n}{j} \sim \binom{n}{\alpha n} \sim \sqrt{\frac{1}{2\pi\alpha(1-\alpha)}} \left(\alpha^{-\alpha}(1-\alpha)^{-(1-\alpha)}\right)^n,$$

using the optimistic (Lanczos) bound above, the time cost is at most polynomial times $\left(\alpha^{-2\alpha}(1-\alpha)^{-2(1-\alpha)}\right)^n$. If $\alpha \approx 0.09$ then $C_{XL/2} \sim 2^{7m/8} \cdot (\text{rational in } m)$. □

This means that XL over $\mathrm{GF}(2)$ *will eventually be a little better than brute-force.*

To show that in fact, decreasing n by fi xing variables does not do much when $q = 2$, we let $m - n$ be 10, 25, and 50, and plot three D_0-vs.-m graphs using Eq. 14:

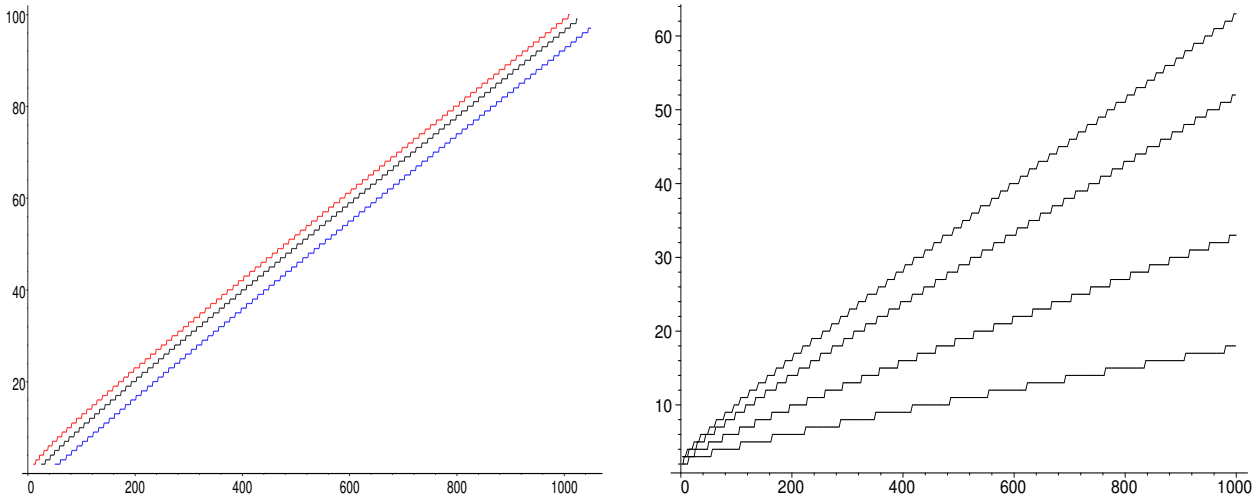
One is hard-pressed to see the difference in the three lines in Fig. 2(a)! For any fi xed $m-n$, D_0/m approaches the same constant as $m \to \infty$. Hence as a speed improvement of XL, XFL/FXL are lacking. Obviously, they may have other useful traits.

Does things change much if we instead make m/n a fi xed ratio? It seems not, see Fig. 2(b): Although the slopes are different, D_0/n still seems to approach a limit.

We are yet to understand all of the interesting behavior in XL, even asymptotically for large m and n, because the coefficients of generating functions are hard enough to estimate, let alone the sign-change point. We leave some for a later work ([24]).

4 XL2 (and XL') over Smaller Fields

One problem with the methods of XL' and XL2 is that they really require general and not sparse matrix methods. However, Strassen-like methods can still be used. Therefore, over $\mathrm{GF}(2)$, XL2 and XL' may show good advantage, a small advantage or no advantage compared to running just XL with well-designed sparse-like algorithms (even



(a) $m - n = 10, 25, 50$ (b) $m/n = \frac{4}{3}, \frac{3}{2}, 2, 3$

Fig. 2. Behavior of D_0 in XL over $\mathrm{GF}(2)$ for fixed $m - n$ and m/n

though we are not sure how sparse matrices mesh with $\mathrm{GF}(2)$), depending on how well everything can be optimized, and the exact comparison is still up in the air.

That said, it is likely possible to inject some extra insight. The authors of [10] seem to consider XL' inferior to XL2, so we aim to have some meaningful discussion about XL2 for both $\mathrm{GF}(2)$ (for which XL2 had originally been designed) and larger fi elds where formerly only heuristics are available, viz. the following observations about XL2. While we believe these to be correct (tested only for small dimensions), the discussions above may not constitute mathematically rigorous proofs.

1. For large q, i.e. $q > D$, every $\mathcal{T}'_j$ is the same, in fact it is exactly $\mathcal{T}^{(D-1)}$. So XL2 can simultanenously apply to any subset of one, two or many variables, because $\mathcal{R}_1 = \mathcal{R}_2 = \cdots = \hat{\mathcal{R}}$, $\mathcal{R}'_1 = \mathcal{R}'_2 = \cdots = \mathcal{R}'$, so by multiplying the C relations $\mathcal{R}'$ by every variable x_i we can maximize the return from XL2, saving a little time.
2. It is not necessarily true that if $I - (T - T') = C > 0$ we will be able to run XL2. The reason is that for any j, all the monomials in $\mathcal{T} \setminus \mathcal{T}'_j$ are at the top degree D. Not all of the I independent equations have those terms.

   Let us illustrate with an example. For large q, we take $m = 11$, $n = 7$, and $D = 3$. Now we have $11 \times (7 + 1) = 88$ equations in XL, all of them independent and $\binom{7+2}{3} = 84$ cubic monomials. It seems as if we should be able to run XL2. Not so, because *only* 77 *of the equations actually have cubic terms.*
3. We can expand on the preceding discussion a little to fi nd when XL2 can be run for larger q. There are $R^{(D)} - R^{(D-1)}$ equations at the top degree. There are usually $R^{(D)} - I^{(D)}$ dependencies, i.e. linear relations between equations, but we need to eliminate the $I^{(D-1)}$ independent equations from those, so there are

$$(R^{(D)} - R^{(D-1)}) - (R^{(D)} - I^{(D)} - I^{(D-1)}) = I^{(D)} - (R^{(D-1)} - I^{(D-1)})$$

independent equations involving the top-degreed monomials, of which there are $T^{(D)} - T^{(D-1)}$. So the condition we seek is: for $q > D$, XL2 operates if:

$$D < T^{(D)} - I^{(D)} < T^{(D-1)} - (R^{(D-1)} - I^{(D-1)}). \tag{19}$$

The term between the parenthesis is our correction to [10] for the large q case. For small q, the behavior is too complex for us to analyze it completely up to now. However, we can do things case-by-case. E.g., we diagnose for the HFE challenge ($m = 80$, $n = 80$, $q = 2$) case, that XL2 runs at $D = 11$, saving only one degree.

4. For $D \geq q > 2$ we can still run XL2. Here, $\mathcal{T}'_j$ generically comprises all monomials of degree less than D (i.e. every monomial in $\mathcal{T}^{(D-1)}$), plus all degree-D monomials where the exponent of x_j is exactly $q - 1$. So

$$T' = [t^D] \frac{(1-t^q)^{n-1}}{(1-t)^n} \left(\frac{t(1-t^q)}{1-t} + t^{q-1} \right).$$

5. Generally speaking, running XL2 on *all* independent variables x_i is equivalent to running XL at one degree higher, regardless of q. How so? Multiply each original XL equation in $\mathcal{R}$ by x_1. This new set of equations (we write $x_1\mathcal{R}$ for short) have only monomials in $x_1\mathcal{T} = \{x_1 p : \ p \in \mathcal{T}\}$. Run an elimination on $x_1\mathcal{R}$ to write every monomial in $x_1\mathcal{T} \setminus \mathcal{T}$ in terms of $x_1\mathcal{T} \cap \mathcal{T} = x_1\mathcal{T}'_1$. These can be bijectively mapped to the equations $x_1\mathcal{R}_1$, and the remaining C equations with only $x_1\mathcal{T}'_1$ monomials correspond similarly to the equations $x_1\mathcal{R}'_1$. Further substitution with relations of x_2 may simplify the equations but does not add new ones.
So, running XL2 with x_i effectively includes equations $x_i\mathcal{R}$ at one degree higher. Running it with every variable will consequently raise the degree of XL by 1.

5 Conclusion

XL is clearly an intriguing idea, one that due to its simplicity has the potential to join Gröbner bases methods ([13, 14]) as a premier equation-solving method. Especially over $\mathrm{GF}(2)$, when the objections of [18] seem inapplicable, XL may and should do well. On the other hand, its theory is still woefully incomplete. Even to implement a solver for large m and n for the simple field of $\mathrm{GF}(2)$ still poses a challenge, due to its space requirements, so we do not know yet the final form of XL. In this, we are reminded of the 2^{63} bytes storage requirement that was suggested in [5]. Clearly some breakthrough in the form of less unwieldy space management techniques needs to be found, as witness the amount of sparse matrix algebra used by Faugère for $F_5/2$ ([13]).

Our misgivings aside, we sincerely hoped to have shed some light on the subject. Generating functions provide a relatively easy way to check if any particular combinations of dimensions will be an operative case for any XL variant. Much remains to be done. Even if behavior of coefficients in generating function can be asymptotically determined, an actual optimization with an obvious space-time tradeoff will still be hard.

Still, we do hope to see a practically useful XL solver at some point.

Acknowledgements

We thank everyone who supported us or made helpful suggestions or comments during the work leading to this manuscript. The first author would especially like to thank (a) his friend and mentor, Dr. Yeong-Nan Yeh of the Institute of Mathematics, Academia Sinica (Taipei) for his support and encouragement; and (b) his ever-tolerant Ping.

References

1. M. Bardet, J.-C. Faugère, and B. Salvy, *Complexity of Gröbner Basis Computations for Regular Overdetermined Systems*, INRIA RR. No. 5049 and private communication.
2. J. R. Bunch and J. E. Hopcroft, *Triangular Factorizations and Inversion by Fast Matrix Multiplication*, Math. Computations, 24 (1974), pp. 231–236.
3. D. Bernstein, *Matrix Inversion Made Diffi cult*, preprint available at `http://cr.yp.to`.
4. N. Courtois, *The Security of Hidden Field Equations (HFE)*, CT-RSA 2001, LNCS V. 2020, pp. 266–281.
5. N. Courtois, *Higher-Order Correlation Attacks, XL Algorithm and Cryptanalysis of Toyocrypt*, ICISC '02, LNCS V. 2587, pp. 182–199.
6. N. Courtois, *Fast Algebraic Attacks on Stream Ciphers with Linear Feedback*, CRYPTO'03, LNCS V. 2729, pp. 176-194.
7. N. Courtois, *Algebraic Attacks over* $\mathrm{GF}(2^k)$*, Cryptanalysis of HFE Challenge 2 and SFLASH*^{v2}, proc. PKC 2004, LNCS V. 2947, pp. 201-217.
8. N. Courtois, A. Klimov, J. Patarin, and A. Shamir, *Effi cient Algorithms for Solving Overdefi ned Systems of Multivariate Polynomial Equations*, EUROCRYPT 2000, LNCS V. 1807, pp. 392–407.
9. N. Courtois and J. Pieprzyk, *Cryptanalysis of Block Ciphers with Overdefi ned Systems of Equations*, ASIACRYPT 2002, LNCS V. 2501, pp. 267–287.
10. N. Courtois and J. Patarin, *About the XL Algorithm over GF(2)*, CT-RSA 2003, LNCS V. 2612, pp. 141–157.
11. C. Diem, *The XL-algorithm and a conjecture from commutative algebra*, preprint to appear ASIACRYPT 2004, and private communication.
12. I. S. Duff, A. M. Erismann, and J. K. Reid, *Direct Methods for Sparse Matrices*, published by Oxford Science Publications, 1986.
13. J.-C. Faugère, *A New Effi cient Algorithm for Computing Gröbner Bases without Reduction to Zero (F5)*, Proceedings of ISSAC 2002, pp. 75-83, ACM Press 2002.
14. 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.
15. R. Fröberg, *An inequality for Hilbert Series of Graded Algebras*, Math. Scand. 56(1985) 117-144.
16. M. Garey and D. Johnson, *Computers and Intractability, A Guide to the Theory of NP-completeness*, Freeman and Co., 1979, p. 251.
17. A. Kipnis and A. Shamir, *Cryptanalysis of the HFE Public Key Cryptosystem by Relinearization*, CRYPTO'99, LNCS V. 1666, pp. 19–30.
18. T. Moh, *On The Method of XL and Its Ineffi ciency Against TTM*, http://eprint.iacr.org/2001/047
19. S. Murphy and M. Robshaw, *Essential Algebraic Structures Within the AES*, CRYPTO 2002, LNCS V. 2442, pp. 1–16.
20. J. Patarin, *Hidden Fields Equations (HFE) and Isomorphisms of Polynomials (IP): Two New Families of Asymmetric Algorithms*, EUROCRYPT'96, LNCS V. 1070, pp. 33–48.
21. R. Stanley, *Enumerative Combinatorics* (2 volumes), Camb. Univ. Press, Cambridge.
22. V. Strassen, *Gaussian Elimination is not Optimal*, Numer. Math. 13 (1969) pp. 354-356.
23. B.-Y. Yang and J.-M. Chen, *All in the XL Family: Theory and Practice*, preprint.
24. B.-Y. Yang and J.-M. Chen, *On Exact and Asymptotic Security Estimates in XL-Related Algebraic Cryptanalysis*, to appear ICICS 2004.