Boolean Ring Cryptographic Equation Solving

Sean Murphy[1], Maura Paterson[2] and Christine Swart[3]

[1] Royal Holloway, University of London, U.K.
[2] Birkbeck, University of London, U.K.
[3] University of Cape Town, South Africa

Abstract. This paper considers multivariate polynomial equation systems over GF(2) that have a small number of solutions. This paper gives a new method `EGHAM2` for solving such systems of equations that uses the properties of the Boolean quotient ring to potentially reduce memory and time complexity relative to existing `XL`-type or Gröbner basis algorithms applied in this setting. This paper also establishes a direct connection between solving such a multivariate polynomial equation system over GF(2), an `MQ` problem, and an instance of the `LPN` problem.
Keywords. `MQ` problem, `XL` algorithm, Gröbner basis, Boolean ring, `LPN` problem.

1 Introduction

This paper considers the `MQ` problem of solving a multivariate nonlinear polynomial equation system over the finite field GF(2), which is an NP-hard problem [24]. The `MQ` problem arises in cryptology in the algebraic cryptanlysis of symmetric primitives [12, 17, 33] and in the analysis of asymmetric schemes based explicitly on this problem [14, 27, 31], an area known as *multivariate cryptography*. In particular, there has been much research on multivariate cryptography, such as [1, 4–6, 8, 11, 15, 16, 18–20, 22, 23, 25, 26, 32]. More recently, a number of multivariate cryptographic schemes have been submitted to the ongoing NIST Post-Quantum Cryptography Standardisation process [30], and the multivariate signature schemes `Rainbow` and `GeMSS` have been selected as a Finalist and an Alternate in this NIST Post-Quantum process.

The contribution of this paper is to develop a method, the `EGHAM2` process, for solving multivariate nonlinear polynomial systems that is specifically adapted for an underlying field of GF(2). For such polynomial systems over GF(2) with small numbers of solutions, the `EGHAM2` process should generally have smaller memory and time complexity than existing `XL`-type and Gröbner basis algorithms. Furthermore, in equation systems over GF(2) where existing `XL`-type and Gröbner basis algorithms are used to produce multivariate polynomials over GF(2) which factorise, the `EGHAM2` process potentially produces many linear expressions which hold with approximate probability $\frac{3}{4}$, so establishing a connection between solving such an equation system over GF(2) and the *Learning Parity with Noise* or `LPN` problem [7].

2 Cryptographic Equation Systems and the Boolean Ring

We consider the problem of finding a solution to the equation system

$$f_1 = \cdots = f_m = 0, \tag{1}$$

where $f_1, \ldots, f_m$ are (without loss of generality) homogeneous polynomials in the multivariate polynomial ring $\mathrm{GF}(2)[x_0, \ldots, x_n]$. We assume the homogeneous system has a small number of solutions, and that they lie in $\mathrm{GF}(2)^{n+1}$. In general, it is also the case that $m \geq n+1$. Such assumptions might reasonably be expected to hold true for systems of equations arising from cryptographic applications where a unique nonzero solution corresponds (for example) to a key that has been used for encryption.

Any element of $\mathrm{GF}(2)$ is fixed by the Frobenius automorphism that sends an element x to x^2. Any point $(x_0, \ldots, x_n) \in \mathrm{GF}(2)^{n+1}$ that is a solution to (1) is therefore also a solution to the (inhomogeneous) polynomial equations $x_i^2 + x_i = 0$ (for $i = 0, \ldots, n$). These are frequently referred to as the *field equations*, and a standard approach is to seek a solution to the (inhomogeneous) equation system

$$f_1 = \cdots = f_m = x_0^2 + x_0 = \cdots = x_n^2 + x_n = 0. \tag{2}$$

We consider an alternative approach to appending the field equations to the original equation system $f_1 = \cdots = f_m = 0$ in the polynomial ring $\mathrm{GF}(2)[x_0, \ldots, x_n]$. Instead, we work in the *Boolean ring* of Definition 1 arising as the quotient ring specified by the ideal generated by these field polynomials.

Definition 1. The *Boolean ring* is the multivariate quotient ring

$$\mathcal{B} = \frac{\mathrm{GF}(2)[x_0, \ldots, x_n]}{\langle x_0^2 + x_0, \ldots, x_n^2 + x_n \rangle}.$$

The canonical ring homomorphism $\Gamma \colon \mathrm{GF}(2)[x_0, \ldots, x_n] \to \mathcal{B}$ or *Boolean mapping* Γ is given by

$$f \mapsto f + \langle x_0^2 + x_0, \ldots, x_n^2 + x_n \rangle. \qquad \square$$

The Boolean ring $\mathcal{B}$ is a principal ideal domain with $z^2 = z$ for all $z \in \mathcal{B}$ [2]. For notational convenience we set $z_i = \Gamma(x_i)$ (for $i = 0, \ldots, n$), and we generally write $g = \Gamma(f)$ for the image of a homogeneous polynomial f and in particular $K = \Gamma(L)$ for the image of a homogeneous linear polynomial L and so on. Thus the Boolean ring $\mathcal{B}$ is a vector space of dimension 2^{n+1} over $\mathrm{GF}(2)$, with the set of all squarefree monomials in the $z_0, z_1, \ldots, z_n$ (including 1) forming a basis. We also let $\mathcal{B}_r$ denote the subspace of $\mathcal{B}$ generated by all such basis monomials of degree at most r. We note that any element of the Boolean ring $\mathcal{B}$ arising as the Boolean image of a homogeneous polynomial has constant term 0.

The Boolean mapping Γ can be applied to each of the polynomials in the equation system $f_1 = \ldots = f_m = 0$ over $\mathrm{GF}(2)$ given by (1) to obtain an equation system in the Boolean ring $\mathcal{B}$ given by

$$\Gamma(f_1) = \cdots = \Gamma(f_m) = 0. \tag{3}$$

Thus equation system (3) can be expressed as $g_1 = \cdots = g_m = 0$ with $g_i = \Gamma(f_i)$. In any case, any element $(x_0, \ldots, x_n) \in \mathrm{GF}(2)^{n+1}$ that is a solution to (1) gives a solution $(z_0, \ldots, z_n) \in \mathcal{B}$ to (3) and vice versa. Our approach in this paper is to seek solutions to the GF(2) system (1) by finding solutions to the corresponding equivalent Boolean system (3).

3 The XL and EGHAM Processes

Many of the proposed approaches for addressing the MQ problem are variants of approaches based on computing Gröbner bases [10], and one such approach is the XL algorithm and its variants [11, 15, 34]. In particular, a geometrically invariant XL approach is considered by [28, 29]. We develop these geometric ideas by giving an improved cryptographic equation solving algorithm (EGHAM2) when the underlying field is GF(2). This improvement is obtained by considering the equation system in the Boolean ring $\mathcal{B}$ rather than in the original polynomial ring $\mathrm{GF}(2)[x_0, \ldots, x_n]$.

3.1 XL-type Algorithms

For our purposes, the approach of the XL algorithm (and variants) can be described in the following way. A homogeneous equation system of degree D is produced from the original equation system $f_1 = \ldots = f_m = 0$ by multiplying the original polynomials $f_1, \ldots, f_m$ by appropriate monomials. Any such resulting polynomial can be represented as a (row) vector of coefficients with respect to a specified basis of monomials of degree D. The vectors corresponding to a basis for the vector space of all such resulting polynomials give a matrix with these vectors as rows known as the *Macaulay matrix*. By considering an appropriate monomial ordering (corresponding to a Macaulay column ordering), Gaussian elimination can be used efficiently to find (if it exists) a bivariate polynomial in two specified variables in the span of this new system of degree D. If such a bivariate polynomial can be found, then it can be potentially factorised into linear factors, one of which gives information about the solution. Such information essentially allows us to remove one variable from the equation system and so on. If no such bivariate polynomial can be found, the process can be repeated by increasing the degree D. The XL algorithm is summarised in Figure 1. A similar description can also be given for Gröbner basis algorithms under an appropriate monomial ordering.

In addition to the number m of polynomials and the number $n' = n + 1$ of variables of the original system, the complexity of the XL algorithm clearly depends fundamentally on the degree D required to find a such a bivariate polynomial. Furthermore, not all of the linear factors of the bivariate polynomial give information about any possible solutions. Determining which linear factors of this bivariate polynomial give information about any possible solutions to the system is a potential further complicating issue in assessing the complexity of the XL algorithm.

1. Consider the system of degree $D \geq 2$ homogeneous polynomials obtained by multiplying each polynomial f_i by the possible monomials of appropriate degree. The resulting system can be expressed in terms of the *Macaulay matrix* $M_{d,m}$ whose columns correspond to the degree D monomials in $\mathrm{GF}(2)[x_0, \ldots, x_n]$ and whose rows correspond to the degree D polynomials in the system. Entries in a given row are the coefficients of the various monomials in the corresponding polynomial.
2. Seek a linear combination of these degree D polynomials that involves only two variables. This can be done by selecting an appropriate ordering for the columns of $M_{d,m}$ then performing Gaussian elimination.
3. Such a homogeneous polynomial in two variables can be factored into linear factors using one of the standard factoring algorithms for univariate polynomials.
4. An appropriate linear factor of this two variable polynomial essentially determines the value of one of the coordinates in the solution. By substituting this value into the original system of equations we can reduce the number of variables by one.
5. By repeating the above steps we hope to find the value of all the coordinates, and hence recover the full solution.

Fig. 1. The `XL` Algorithm

3.2 The `EGHAM` Process

The basic `XL` algorithm is not geometrically invariant as a simple linear change of co-ordinates can greatly change the complexity. This motivated the development of geometrically invariant forms of the `XL` algorithm [28, 29]. The `EGHAM` (Even Geometric Heuristic Algorithmic Method) process [29] is such a geometrically invariant `XL`-type algorithm specially designed for equation systems where the underlying field has characteristic 2. The fundamental concept of the `EGHAM` process is the geometrically invariant generalisation of the homogeneous bivariate polynomial to the Rank-2 Product Polynomial, which is given in its GF(2) formulation in Definition 2. The property of a Rank-2 Product Polynomial giving rise to this terminology is then given in Lemma 1 (proved in [28]). The development of such an approach then yields the $\mathcal{LS}$-Criterion [29] of Definition 3.

Definition 2. *A Rank-2 Product Polynomial* of degree D is a homogeneous polynomial of the form $\prod_{i=1}^{D}(\theta_i' L' + \theta_i'' L'') \in \mathrm{GF}(2)[x_0, ..., x_n]$, where L' and L'' are homogeneous linear polynomials over GF(2) and θ_i' and θ_i'' are constants in some extension field of GF(2). □

Lemma 1. The matrix C_f of formal partial derivatives of a Rank-2 Product Polynomial f has rank at most 2. □

Definition 3. Let $W_D \subset \mathrm{GF}(2)[x_0, \ldots, x_n]$ denote the space of homogeneous polynomials of degree D over GF(2). A homogeneous polynomial $f \in W_D$ (for $D > 0$) satisfies the *$\mathcal{LS}$-Criterion* if f is an element of

- either the $\mathcal{L}^2\mathcal{S}$ subspace $\left\langle x_i x_j \mathbf{x}^2 \,\middle|\, \mathbf{x} \in W_{\frac{1}{2}(D-2)} \right\rangle$ when D is even
- or the $\mathcal{L}^1\mathcal{S}$ subspace $\left\langle x_i \mathbf{x}^2 \,\middle|\, \mathbf{x} \in W_{\frac{1}{2}(D-1)} \right\rangle$ when D is odd. □

In particular, Lemma 2 (proved in [29]) shows that the $\mathcal{LS}$-Criterion categorises the Rank-2 Product Polynomials. Example 1 then illustrates Lemma 2 with a Rank-2 Produduct polynomial that satisfies the $\mathcal{LS}$-Criterion.

Lemma 2. A Rank-2 Product Polynomial satisfies the $\mathcal{LS}$-Criterion. □

Example 1. We consider the homogenous polynomial f of degree 4 in the polynomial ring GF(2)$[x_0, x_1, x_2]$ given by

$$\begin{aligned} f &= x_0^4 + x_0x_1^3 + x_0x_1^2x_2 + x_1^3x_2 + x_0x_1x_2^2 + x_1^2x_2^2 + x_0x_2^3 + x_1x_2^3 \\ &= L'L''(L' + \omega L'')(L' + \omega^2 L''), \end{aligned}$$

where $L' = x_0 + x_1$, $L'' = x_0 + x_2$ and ω is a root of $y^2 + y + 1 = 0$ over GF(2). The product form for f shows that f is a Rank-2 Product Polynomial and the monomials of f show that f satisfies the $\mathcal{LS}$-Criterion. Furthermore, the partial derivatives matrix C_f is given with respect to the lexicographic monomial ordering $(x_0^3, x_0^2x_1, x_0^2x_2, x_0x_1^2, x_0x_1x_2, x_0x_2^2, x_1^3, x_1^2x_2, x_1x_2^2, x_2^3)$ by

$$C_f = \begin{pmatrix} 0\,0\,0\,0\,0\,0\,1\,1\,1\,1 \\ 0\,0\,0\,1\,0\,1\,0\,1\,0\,1 \\ 0\,0\,0\,1\,0\,1\,1\,0\,1\,0 \end{pmatrix}.$$

This partial derivatives matrix C_f of f has rank 2 over GF(2). □

Lemma 2 gives rise to the `EGHAM` process of [29], which we now outline. The $\mathcal{L}^2\mathcal{S}$ subspace or the $\mathcal{L}^1\mathcal{S}$ subspace have dimension in general far smaller than the subspace generated by the homogeneous degree D polynomials under consideration. Thus taking the intersection of this subspace generated by these polynomials with the $\mathcal{L}^2\mathcal{S}$ subspace or the $\mathcal{L}^1\mathcal{S}$ subspace allows us to use the $\mathcal{LS}$-Criterion as a highly efficient filter to vastly reduce the number of polynomials under consideration. Suppose therefore (without loss of generality) that $f_1, \ldots, f_m$ are the homogeneous polynomials of degree D in an `XL`-type process obtained after filtering using the $\mathcal{LS}$-Criterion and that $f_1, \ldots, f_m$ form a basis for this resulting subspace. We can associate an appropriate partial derivatives matrix C_{f_i} with each polynomial f_i $(i = 1, \ldots, m)$, so any polynomial $\sum_{i=1}^m \lambda_i f_i$ in the span of $f_1, \ldots, f_m$ has corresponding partial derivatives matrix $\sum_{i=1}^m \lambda_i C_{f_i}$. Lemma 1 shows that a Rank-2 Product Polynomial in the span of $f_1, \ldots, f_m$ has corresponding partial derivatives matrix of rank 2, that is to say we would require all 3×3 subdeterminants of $\sum_{i=1}^m \lambda_i C_{f_i}$ to be 0. This gives rise to a cubic equation system in $\lambda_1, \ldots, \lambda_m$ whose solutions correspond to Rank-2 Product Polynomials and hence potentially to information about the solution to the original equation system.

The `EGHAM` process is summarised in Figure 2, and there are polynomial systems for which the `EGHAM` process works with a far lower degree of D than is required by `XL` or standard Gröbner basis algorithms [29]. However, having to solve a cubic system in $\lambda_1, \ldots, \lambda_m$ is not ideal, and this is one of the issues that we seek to address with the `EGHAM2` process.

1. Consider the system of degree $D \geq 2$ homogeneous polynomials obtained by multiplying each polynomial f_i by the possible monomials of appropriate degree.
2. Apply the $\mathcal{LS}$-Criterion, restricting attention to either the $\mathcal{L}^2\mathcal{S}$ or $\mathcal{L}^1\mathcal{S}$ subspace as required, thus reducing the dimension of the problem.
3. Find a Rank-2 Product Polynomial in the $\mathcal{L}^2\mathcal{S}$ or $\mathcal{L}^1\mathcal{S}$ subspace. The approach suggested in [29] requires the solution of a system of cubic equations.
4. The Rank-2 Product Polynomial can be factored, and the appropriate substitution then reduces the number of variables by one.
5. This process is repeated until the desired solution is found.

Fig. 2. A Summary of the `EGHAM` Process

3.3 A Boolean view of the `EGHAM` process

We now consider how various aspects of the `EGHAM` process are affected when we move to working directly in the Boolean ring $\mathcal{B}$. In particular, we consider those polynomials whose Boolean image has degree at most 2, giving the $\mathcal{Q}$-Criterion of Definition 4. Lemma 3 then shows that Rank-2 Product Polynomials satisfy this $\mathcal{Q}$-Criterion.

Definition 4. A homogeneous polynomial $f \in W_D$ $(D > 0)$ of degree D satisfies the *Quadratic Criterion* or $\mathcal{Q}$-Criterion if $g = \Gamma(f) \in \mathcal{B}_2$, that is to say the image of f under the Boolean mapping Γ is quadratic or linear or 0. □

Lemma 3. The image of a Rank-2 Product Polynomial under the Boolean mapping Γ is either a linear Boolean element or is a quadratic Boolean element of the form $K'K'' + K' + K''$ for linear Boolean elements K' and K''. Thus a Rank-2 Product Polynomial satisfies the $\mathcal{Q}$-Criterion.

Proof. Let $f \in \mathrm{GF}(2)[x_0, \ldots, x_n]$ be a Rank-2 Product Polynomial of degree $D \geq 2$ whose factorisation over an extension of GF(2) is $f = \prod_{i=1}^{D}(\theta_i' L' + \theta_i'' L'')$ for some homogeneous linear polynomials L' and L'' over GF(2), so f can be expressed as

$$f = {L'}^D + \sum_{i=1}^{D-1} c_i {L'}^i {L''}^{D-i} + {L''}^D$$

with $c_1, \ldots c_{D-1} \in \mathrm{GF}(2)$. The Boolean image of f is therefore given by

$$\begin{aligned} g = \Gamma(f) &= \Gamma(L')^D + \sum_{i=1}^{D-1} c_i \Gamma(L')^i \Gamma(L'')^{D-i} + \Gamma(L'')^D \\ &= K' + (c_1 + \ldots + c_{D-1}) K'K'' + K'', \end{aligned}$$

where $K' = \Gamma(L')$ and $K'' = \Gamma(L'')$ are linear Boolean elements. Thus $g = K' + K''$ if $c_1 + \ldots + c_{D-1} = 0$ and $g = K' + K'K'' + K''$ if $c_1 + \ldots + c_{D-1} = 1$. In either case, the image of a Rank-2 Product Polynomial f under Γ has degree at most 2 and so f satisfies the $\mathcal{Q}$-Criterion. □

1. Given a set of homogeneous polynomials over GF(2), obtain the corresponding image set of Boolean elements in $\mathcal{B}$ by using the Boolean mapping Γ.
2. From this set of Boolean elements of $\mathcal{B}$, find elements in the span satisfying the $\mathcal{Q}$-Criterion. Such a reduced set of quadratic Boolean equations can be found by taking the intersection of the original Boolean elements with $\mathcal{B}_2$.
3. Find elements in the span of this reduced set of quadratic Boolean elements satisfying the $\mathcal{R}_2$-Criterion by using the Kernel Method or otherwise. Hence find Boolean linear expressions which hold with probability approximately $\frac{3}{4}$.
4. Express these probabilistic Boolean linear expressions as a Learning Parity with Noise (`LPN`) Problem
   (a) Attempt to solve this `LPN` problem using the `BKW` algorithm or otherwise.
   (b) If there are not sufficient Boolean linear expressions to solve this `LPN` problem, then the original equation system can be expanded by multiplying elements by monomials and the process repeated.

Fig. 3. Overview of the `EGHAM2` Process

Example 2. The homogeneous Rank-2 Product Polynomial of Example 1 given by $f = x_0^4 + x_0x_1^3 + x_0x_1^2x_2 + x_1^3x_2 + x_0x_1x_2^2 + x_1^2x_2^2 + x_0x_2^3 + x_1x_2^3$ satisfies the $\mathcal{LS}$-Criterion. The image

$$g = \Gamma(f) = z_0 + z_0z_1 + z_0z_2 + z_1z_2$$

of f in the Boolean ring $\mathcal{B}$ consists only of linear and quadratic terms, so f satisfies the $\mathcal{Q}$-Criterion. □

The $\mathcal{Q}$-Criterion gives a further highly restrictive condition for a Rank-2 Product Polynomial in the Boolean case. For example, $x_0^3x_1^4x_2^5$ satisfies the $\mathcal{LS}$-Criterion, but its image $\Gamma(x_0^3x_1^4x_2^5) = z_0z_1z_2$ under Γ does not satisfy the $\mathcal{Q}$-Criterion. This suggests that a development of the `EGHAM` process directly focussed on the Boolean ring $\mathcal{B}$ and the $\mathcal{Q}$-Criterion offers the potential for substantial performance improvements in identifying Rank-2 Product Polynomials.

4 A Boolean `EGHAM` process: `EGHAM2`

We now give a version of the `EGHAM` process that is adapted to the Boolean ring $\mathcal{B}$ based on the ideas of Section 3.3, and we term the resulting Boolean process the `EGHAM2` (Even Geometric Heuristic Algorithmic Method for GF(2)) process. We give a high-level view of this `EGHAM2` process in Figure 3, and we discuss issues relating to this `EGHAM2` process in this Section. However, we note as motivation for this process that the $\mathcal{Q}$-Criterion generally gives a very much smaller set of quadratic elements than the original set of Boolean elements, which is obviously much simpler and more efficient to handle. We also note that the `EGHAM2` process generates probabilistic linear expressions for the solution, so developing a direct relationship between the `MQ` problem and the `LPN` problem

4.1 The Kernel of the Boolean Mapping

The ideal $\langle x_0^2 + x_0, \ldots, x_n^2 + x_n \rangle$ generated by the field equations is by definition the kernel of the Boolean mapping Γ, and so $\ker(\Gamma)$ plays a critical role in the development of the `EGHAM2` process. In particular, this Boolean mapping $\ker(\Gamma)$ allows us to extend the ideas of Section 3 to certain polynomials that are not Rank-2 Product Polynomials.

The ideas underlying the use of this kernel can be illustrated by considering the polynomial $f_0 = x_i^2 x_j + x_j^2 x_k = x_j(x_i^2 + x_j x_k)$. The polynomial f_0 factorises, but is not itself a Rank-2 Product Polynomial. We do observe however that $\Gamma(f_0) = z_i z_j + z_j z_k = z_j(z_i + z_k)$ is a quadratic element which does factorise. Such a quadratic element factorisation occurs as f_0 differs from a Rank-2 Product Polynomial by an element of $\ker(\Gamma)$. In this case we have $x_j x_k^2 + x_j^2 x_k \in \ker(\Gamma)$, which gives

$$f_0 + (x_j x_k^2 + x_j^2 x_k) = x_i^2 x_j + x_j x_k^2 = x_j(x_i + x_k)^2,$$

so $f_0 + (x_j x_k^2 + x_j^2 x_k)$ is a Rank-2 Product Polynomial satisfying

$$g_0 = \Gamma(f_0) = \Gamma\left(f_0 + (x_j x_k^2 + x_j^2 x_k)\right) = z_j(z_i + z_k).$$

For this example, the application of the Boolean mapping Γ has shown us that the ideal generated by f_0 and the elements of $\ker(\Gamma)$ does contain a Rank-2 Product Polynomial, and has allowed us to find its image.

Applying the Boolean mapping Γ to a single polynomial gives an image that essentially gives us information about its "most useful" preimage, and the same notion can be extended to systems of polynomials. Adding polynomials in the Boolean mapping kernel $\ker(\Gamma)$ to the polynomials defining the set of equations we wish to solve does not affect the solutions over GF(2). However, adding such "kernel polynomials" can significantly lower the smallest degree D for which the `EGHAM` process succeeds. All polynomials obtained in this way have the same images under Γ, and so a process based on the resulting Boolean equation system, such as the `EGHAM2` process, works for the lowest degree D that succeeds for any of these possible preimages of this system. This idea is illustrated by the following Examples which consider two polynomial equation systems that have the same image under Γ. Example 3 gives a homogeneous cubic polynomial equation system that yields a direct factorisation, so potentially giving a solution to the equation system, only using these cubic polynomials, whilst Example 4 gives a similar polynomial equation system that does not give such a factorisation using cubic polynomials. However, Example 5 shows that the Boolean image of these polynomial equation systems yields a factorisation in both cases.

Example 3. We consider twelve homogenous polynomials $f_1, \ldots, f_{12}$ of degree 3 in the polynomial ring GF(2)$[x_0, x_1, x_2, x_3, x_4, x_5]$ given by

001 003 005 011 013 014 024 034 113 122 124 133 134 144 223 234 235 255 333 335 344 345 355 445 555
002 005 011 012 013 014 022 025 033 034 044 045 112 114 115 133 144 145 222 223 224 235 244 245 255 335 345 444 445
001 002 003 004 005 011 023 034 044 045 055 111 112 113 114 115 123 125 134 145 222 223 224 233 244 245 255 335 555
003 011 012 015 025 044 111 113 114 115 124 125 133 134 145 224 233 234 235 244 245 255 333 345 455 555
000 001 003 015 024 025 045 112 122 124 134 135 233 235 255 333 334 355 455 555
000 001 002 003 005 014 022 023 024 025 044 045 055 112 114 123 144 223 224 234 245 333 334 335 344 345 355 455 555
002 003 004 013 014 022 024 025 033 034 035 044 055 113 114 115 122 124 125 133 135 145 223 224 225 345 355 445 555
001 002 003 005 012 013 023 034 035 045 055 111 114 125 135 225 233 234 244 245 255 334 335 345 355 444 445 455
001 003 005 011 012 013 014 015 022 023 034 045 112 115 124 125 133 135 145 155 222 224 225 233 234 235 333 334 355 444 455
001 004 011 012 013 015 022 023 025 033 111 115 123 134 144 145 155 225 234 235 333 334 344 355 444 445 555
000 001 003 004 005 011 013 022 023 024 033 035 045 113 115 123 145 222 223 234 244 245 333 334 335 344
000 001 002 004 005 011 013 014 015 033 044 111 112 122 123 124 125 133 134 144 224 225 233 244 245 333 334 335 344 355 555

The notation abc denotes the monomial $x_a x_b x_c$ and addition signs are omitted, so for example 000 011 123 would denote the polynomial $x_0^3 + x_0 x_1^2 + x_1 x_2 x_3$, and each line gives a single polynomial. The equation system $f_1 = \ldots = f_{12} = 0$ has the unique nonzero solution $x^* = (1, 1, 0, 0, 1, 0)$. To find this solution using the `EGHAM` process we apply the $\mathcal{LS}$-Criterion, when we obtain the single polynomial

$$x_0^3 + x_0^2 x_2 + x_0^2 x_3 + x_0 x_1^2 + x_0 x_3^2 + x_0 x_5^2 + x_1^2 x_2 + x_1^2 x_3 + x_2 x_3^2 + x_2 x_5^2 + x_3^3 + x_3 x_5^2$$

in the span of the above system. This $\mathcal{LS}$-Criterion polynomial factorises as

$$(x_0 + x_2 + x_3)(x_0 + x_1 + x_3 + x_5)^2,$$

which shows that this polynomial is a Rank-2 Product polynomial. At least one of these linear factors evaluated at the solution is 0, and so an appropriate substitution can remove one of the variables from the system to give a simpler polynomial equation system. □

Example 4. We consider twelve homogenous polynomials $f'_1, \ldots, f'_{12}$ of degree 3 in the polynomial ring GF(2)$[x_0, x_1, x_2, x_3, x_4, x_5]$ given by

001 003 005 011 013 014 024 034 115 122 124 134 144 155 223 225 234 235 333 334 345 455 555
001 012 013 014 025 033 034 044 045 055 112 113 114 115 144 145 222 223 225 235 245 335 345 444 455
004 005 022 023 033 034 044 045 055 111 112 115 123 125 133 134 144 145 222 223 233 245 255 355 555
001 005 012 015 025 033 044 055 111 112 114 122 124 125 134 145 155 233 234 235 245 255 333 334 335 344 345 355 445 555
000 002 003 011 015 022 024 025 045 113 114 115 124 133 134 135 144 155 224 225 233 235 244 333 335 344 445 555
000 001 004 014 023 024 025 033 045 114 122 123 144 223 224 225 234 245 255 333 334 344 345 455 555
004 013 014 024 025 034 035 044 055 113 114 122 124 125 133 135 145 155 224 225 233 345 355 445 555
003 005 011 012 013 022 023 034 035 045 055 111 114 125 135 224 225 233 234 245 255 334 335 345 355 444
001 003 005 011 012 013 014 015 022 023 034 045 112 113 124 125 135 145 222 223 225 234 235 244 333 344 355 444 445
001 002 005 011 012 013 015 023 025 033 044 055 111 112 114 115 122 123 134 145 155 224 225 234 235 244 333 335 444 455 555
000 002 003 004 005 013 023 024 033 035 045 114 115 123 133 144 145 222 223 234 244 245 333 335
000 001 002 003 004 005 011 013 014 015 044 111 113 114 123 124 125 134 223 245 255 333 334 344 555

The equation system $f'_1 = \ldots = f'_{12} = 0$ has the unique nonzero solution $(1, 1, 0, 0, 1, 0)$, as in Example 3. Applying the $\mathcal{LS}$-Criterion to this equation system gives the single polynomial

$$\begin{aligned} &x_0^3 + x_0^2 x_2 + x_0^2 x_5 + x_0 x_1^2 + x_1^2 x_2 + x_1^2 x_3 + x_1^2 x_4 + x_1^2 x_5 \\ &+ x_1 x_4^2 + x_1 x_5^2 + x_2^2 x_3 + x_2^2 x_5 + x_3^3 + x_3 x_5^2 + x_4^2 x_5 + x_4 x_5^2. \end{aligned}$$

in the span of the above system. This polynomial is absolutely irreducible over GF(2) and so is not a Rank-2 Product polynomial. This means that there is no degree three polynomial in the ideal generated by these polynomials that is a Rank-2 Product polynomial. We therefore have to generate a higher degree system from this cubic system for the `EGHAM` process to succeed. □

Example 5. Both of the polynomial systems of Example 3 and 4 are homogeneous cubic systems of 12 polynomials in 6 variables. However, these two systems have a common image under the Boolean mapping Γ given by the following Boolean element system

```
013 014 024 034  03  05 124  12 134  14 234 235  23  25 345  34   3  45   5
012 013 014  01 025 034  03 045  04  05  12  13 145  15 235  23 245  25   2 345 35 45  4
023  02 034  03 045 123 125  12 134  13 145  14  15   1 245  25   2  35   5
012 015  01 025  03  04 124 125 134 145  14  15   1 234 235  23 245  25 345   3  45  5
015  01 024 025  03 045   0 124 134 135 235  23  25  34  35   3  45   5
014  01 023 024 025  03 045  04   0 123  12 234  23 245  24 345   3  45   5
013 014 024 025 034 035  05 124 125  12 135 145  14  15  23  24  25 345  35  45  5
012 013  01 023  02 034 035  03 045 125 135  14   1 234  23 245  24 345  34   4
012 013 014 015 023  02 034  03 045  05 124 125  12 135  13 145 234 235  23  24 25 2 34 35 3 45 4
012 013 015 023 025  02  03  04 123 134 145  14   1 234 235  25  35   3  45   4  5
013 023 024  02 035 045  04  05   0 123  13 145  15 234  23 245  24   2  35   3
013 014 015  02  03  05   0 123 124 125 134  13  14   1  23 245  25   3   5
```

If we apply the $\mathcal{Q}$-Criterion to this common image of the polynomial systems of Example 3 and 4 we obtain the Boolean element

$$z_0z_1 + z_0z_2 + z_0z_5 + z_0 + z_1z_2 + z_1z_3 + z_2z_3 + z_2z_5 + z_3z_5 + z_3$$

in the span of the above Boolean elements, which factorises to give

$$(z_0 + z_2 + z_3)(z_0 + z_1 + z_3 + z_5).$$

This Boolean factorisation is the image under the Boolean mapping Γ of the factorisation of Example 3. □

The fundamental point made by these Examples is that formally mapping polynomials in the polynomial ring over GF(2) to elements of the Boolean ring $\mathcal{B}$ allows us to find potentially useful "Boolean factorisations" in the span of a polynomial system that are not generally found by an `EGHAM` process in the polynomial ring. We can use such a Boolean factorisation to give trial substitutions of variables for solving the original polynomial system

4.2 The $\mathcal{R}_2$-Criterion for a Quadratic Boolean Element

The application of the $\mathcal{Q}$-Criterion leads us to find quadratic elements of the Boolean ring. However, we can associate a quadratic element of the Boolean ring $\mathcal{B}_2$ with a matrix essentially given by its partial derivatives, namely the ∂-matrix of Definition 5, a symmetric $(n+1) \times (n+1)$ matrix over GF(2). Our analysis of Rank-2 Product Polynomials proceeds by considering such ∂-matrices for quadratic Boolean elements arising in the span of the image of the polynomial equation system under the Boolean mapping Γ. In particular, we consider the $\mathcal{R}_2$-Criterion of Definition 6, as the subsequent Lemma 4 indicates how to use this $\mathcal{R}_2$-Criterion to locate images of Rank-2 Product Polynomials.

Definition 5. A quadratic element $g = \sum_{i=1}^{n} \sum_{j=0}^{i-1} a_{ij} z_i z_j + \sum_{i=1}^{n} a_{ii} z_i + a \in \mathcal{B}_2$ has symmetric $(n+1) \times (n+1)$ *∂-matrix* ∂g given by $(\partial g)_{ij} = (\partial g)_{ji} = a_{ij}$ for $j < i$ with 0-diagonal $(\partial g)_{ii} = 0$. □

Definition 6. A quadratic Boolean element $g \in \mathcal{B}_2$ satisfies the $\mathcal{R}_2$-*Criterion* if the ∂-matrix ∂g of g has rank at most 2. □

Lemma 4. A Rank-2 Product Polynomial f has an image $g = \Gamma(f)$ under the Boolean mapping Γ which satisfies the $\mathcal{R}_2$-Criterion. □

Proof. If $g = \Gamma(f) \in \mathcal{B}_1$ then $\partial g = \partial\Gamma(f) = 0$. Lemma 3 shows that the remaining possibility for a Rank-2 Product Polynomial f is that

$$g = \Gamma(f) = K'K'' + K' + K''$$

for images $K' = \sum_{i=0}^{n} b'_i z_i$ and $K'' = \sum_{j=0}^{n} b''_j z_j$ of two homogeneous linear polynomials. In this case we have

$$g = K'K'' + K' + K'' = \sum_{i=1}^{n} \sum_{j=0}^{i-1} (b'_i b''_j + b''_i b'_j) z_i z_j + \sum_{i=0}^{n} (b'_i b''_i + b'_i + b''_i) z_i.$$

with the corresponding ∂-matrix ∂g given by $(\partial g)_{ij} = b'_i b''_j + b''_i b'_j$ for $i \neq j$ and $(\partial g)_{ii} = 0$. If we let $b' = (b'_0, \ldots, b'_n)^T$ and $b'' = (b''_0, \ldots, b''_n)^T$ be the column vectors of coefficients of K' and K'', then the ∂-matrix of g is given by

$$\partial g = b'b''^T + b''b'^T.$$

The ∂-matrix ∂g of g is the sum of two matrices $b'b''^T$ and $b''b'^T$ of rank 1, so has rank at most 2. Thus g satisfies the $\mathcal{R}_2$-Criterion. □

Example 6. Example 2 shows that $g = z_0 + z_0z_1 + z_0z_2 + z_1z_2$ is the Boolean image of $f = x_0^4 + x_0x_1^3 + x_0x_1^2x_2 + x_1^3x_2 + x_0x_1x_2^2 + x_1^2x_2^2 + x_0x_2^3 + x_1x_2^3$ of Example 1. This Boolean quadratic element g has ∂-matrix

$$\partial g = \begin{pmatrix} 0\,1\,1 \\ 1\,0\,1 \\ 1\,1\,0 \end{pmatrix},$$

a matrix of rank 2 over GF(2). Thus $g = \Gamma(f)$, the Boolean image of the Rank-2 Product Polynomial f, satisfies the $\mathcal{R}_2$-Criterion.

4.3 Finding Quadratic Elements satisfying the $\mathcal{R}_2$-Criterion

Suppose that $g_1, \ldots, g_m \in \mathcal{B}_2$ are quadratic Boolean elements, such as might be obtained by applying the $\mathcal{Q}$-Criterion to some larger original polynomial equation system, where we assume that there are significantly more resulting quadratic Boolean elements than variables, so $m \gg n$. We consider how to find an element $g = \sum_{i=1}^{m} \lambda_i g_i$ in their span satisfying the $\mathcal{R}_2$-Criterion. The *Kernel Method* [13, 21] is a method to find a matrix of low rank within the span of a set of matrices. We can use this Kernel Method to attempt to find a quadratic Boolean element $g = \sum_{i=1}^{m} \lambda_i g_i$ satisfying the $\mathcal{R}_2$-Criterion in the span of $g_1, \ldots, g_m$ by

finding a $(n+1) \times (n+1)$ matrix in the span of the corresponding ∂-matrices $\partial g_1, \ldots, \partial g_m$ such that

$$\partial g = \sum_{i=1}^{m} \lambda_i \, \partial g_i \quad \text{has rank 2.}$$

A randomly chosen column vector v_1 of dimension $n+1$ lies in the kernel $\ker(\partial g)$ of a matrix ∂g of rank 2 with probability $\frac{1}{4}$, and so

$$v_1^T \left(\sum_{j=1}^{m} \lambda_j \partial g_j \right) = \sum_{j=1}^{m} \lambda_j \left(v_1^T \, \partial g_j \right) = 0 \quad \text{with probability } \tfrac{1}{4}.$$

Any coefficient vector $(\lambda_1, \ldots, \lambda_m)$ satisfying $\sum_{j=1}^{m} \lambda_j \left(v_1^T \, \partial g_j \right) = 0$ lies in the (left) kernel of the $m \times (n+1)$ matrix

$$\Lambda^{(v_1)} = \begin{pmatrix} \hline v_1^T \, \partial g_1 \\ \hline \vdots \\ \hline v_1^T \, \partial g_m \\ \hline \end{pmatrix}.$$

However, $\ker\left(\Lambda^{(v_1)}\right)$ is typically a subspace of large dimension as $m \gg n$, so whilst $\sum_{j=1}^{m} \lambda_j \, \partial g_j$ is not generally a matrix of rank 2 or less for $(\lambda_1, \ldots, \lambda_m)$ in the kernel of $\Lambda^{(v_1)}$, this kernel typically gives rise to many matrices $\sum_{j=1}^{m} \lambda_j \, \partial g_j$ that are of rank 2 or less. We can repeat this process for l further randomly chosen vectors $v_2, \ldots, v_l$ and determine $\ker\left(\Lambda^{(v_1)}\right) \bigcap \ldots \bigcap \ker\left(\Lambda^{(v_l)}\right)$. Thus we can determine coefficient vectors $(\lambda_1, \ldots, \lambda_m)$ that could potentially give rise to a matrix $\sum_{j=1}^{m} \lambda_j \, \partial g_j$ of rank 2 or less by determining the (left) kernel of an $m \times l(n+1)$ matrix, that is to say by solving

$$(\lambda_1, \ldots, \lambda_m) \left(\Lambda^{(v_1)} \middle| \ldots \middle| \Lambda^{(v_l)} \right) = (\lambda_1, \ldots, \lambda_m) \left(\begin{array}{c|c|c} v_1^T \partial g_1 & \ldots & v_l^T \partial g_1 \\ \hline \vdots & \ddots & \vdots \\ \hline v_1^T \partial g_m & \ldots & v_l^T \partial g_m \end{array} \right) = 0.$$

In summary, the Kernel Method can be used to find $(\lambda_1, \ldots, \lambda_m)$ such that the matrix $\partial g = \sum_{i=1}^{m} \lambda_i \partial g_i$ is a candidate to be a matrix of rank 2, corresponding to quadratic Boolean element $g = \sum_{i=1}^{m} \lambda_i g_i$ satisfying the $\mathcal{R}_2$-Criterion. We note that we can generate the matrices $\partial g_1, \ldots, \partial g_m$ by an echelon-like process, so they are themselves likely to be matrices of low rank, meaning that we are likely to find matrices of the form $\sum_{i=1}^{m} \lambda_i \partial g_i$ of rank 2. Furthermore, we choose l such that the required kernel, corresponding to possible candidates $(\lambda_1, \ldots, \lambda_m)$, is not too large. Example 8 of Section 4.7 contains a brief discussion about a process for determining l, and we note that an appropriate size for l can easily be determined empirically. By repeating this process, the Kernel Method potentially allows us to generate many such quadratic Boolean elements $g = \sum_{i=1}^{m} \lambda_i g_i$ in the span of $g_1, \ldots, g_m$ satisfying the $\mathcal{R}_2$-Criterion.

4.4 Probabilistic Linear Expressions

Lemma 4 shows that the $\mathcal{R}_2$-Criterion provides a useful filter for determining whether a polynomial is a Rank-2 Product Polynomial. It is possible for the images of other polynomials which are not Rank-2 Product Polynomials to satisfy the $\mathcal{R}_2$-Criterion, for example $x_0^2 + x_1x_2$ has image $\Gamma(x_0^2 + x_1x_2) = z_0 + z_1z_2$ with ∂-matrix of rank 2. This issue arises as the ∂-matrix depends only on the quadratic coefficients. However, Lemma 5 gives a decomposition for quadratic elements satisfying the $\mathcal{R}_2$-Criterion, which yields probabilistic Boolean linear expressions.

Lemma 5. Suppose that a nontrivial quadratic element $g \in \mathcal{B}_2$ (with constant term 0) satisfies the $\mathcal{R}_2$-Criterion, then there exist homogeneous linear elements $K, K', K'' \in \mathcal{B}_1$ such that $g = K'K'' + K$. If g takes the value 0 and K is not identically 0, then K takes the value 0 with probability approximately $\frac{3}{4}$. □

Proof. A nontrivial ∂-matrix cannot have rank 1 as it is symmetric. Thus suppose that b'^T and b''^T are two linearly independent rows of the ∂-matrix ∂g, so b'^T and b''^T form a basis for the rowspace of ∂g and $\partial g = b'b''^T + b''b'^T$ as ∂g is symmetric. If we define the linear elements $K' = \sum_{i=0}^{n} b'_i z_i$ and $K'' = \sum_{i=0}^{n} b''_i z_i$, then clearly $\partial g = \partial(K'K'')$, and so g and $K'K''$ can differ only in their linear terms. Thus we can write $g = K'K''+K$ for some linear Boolean element $K \in \mathcal{B}_1$. Furthermore, if $g = K'K'' + K$ takes the value 0 and K is not identically 0, then $(K, K', K'') \in \{(0,0,0), (0,0,1), (0,1,0), (1,1,1)\}$, and in three out of the four cases K takes the value 0. As we exclude the trivial $z = 0$ solution, then K takes the value 0 with probability $\frac{3}{4} - 2^{-n} \approx \frac{3}{4}$. □

Lemma 5 indicates how to use the ∂-matrix ∂g of a Boolean element g satisfying the $\mathcal{R}_2$-Criterion to find probabilistic linear Boolean expressions. If $\partial g \neq 0$, then ∂g is a matrix of rank 2 over GF(2), so has two linearly independent rows, corresponding to the distinct linear Boolean elements K' and K''. We note that any third distinct nonzero row of ∂g corresponds to the linear Boolean element $K' + K''$. If we then construct the three linear Boolean elements

$$L = g + K'K'', \quad L' = g + K'(K' + K'') \quad \text{and} \quad L'' = g + K''(K' + K''),$$

then Lemma 5 shows that L, L' and L'' take the value 0 with probability $\frac{3}{4}$ if they are not identically 0, though we note that L, L' and L'' are correlated random variables. We also note that $L + L' + L''$ takes the value 0 with approximate probability $\frac{1}{4}$.

Example 7. Consider the quadratic Boolean element

$$g = z_0+z_0z_1+z_2+z_1z_2+z_0z_3+z_1z_3+z_2z_3+z_0z_4+z_2z_4+z_3z_4+z_1z_5+z_3z_5+z_4z_5$$

with six variables, which has ∂-matrix over GF(2) given by

$$\partial g = \begin{pmatrix} 0\,1\,0\,1\,1\,0 \\ 1\,0\,1\,1\,0\,1 \\ 0\,1\,0\,1\,1\,0 \\ 1\,1\,1\,0\,1\,1 \\ 1\,0\,1\,1\,0\,1 \\ 0\,1\,0\,1\,1\,0 \end{pmatrix}$$

of rank 2, so g satisfies the $\mathcal{R}_2$-Criterion. The first and second rows of ∂g correspond to the linear Boolean elements $K' = z_1+z_3+z_4$ and $K'' = z_0+z_2+z_3+z_5$, so we obtain (corresponding to fourth row) $K' + K'' = z_0 + z_1 + z_2 + z_4 + z_5$. Thus we can obtain the three linear Boolean elements

$$\begin{aligned} L &= g + (z_1 + z_3 + z_4)(z_0 + z_2 + z_3 + z_5) = z_0 + z_2 + z_3, \\ L' &= g + (z_1 + z_3 + z_4)(z_0 + z_1 + z_2 + z_4 + z_5) = z_0 + z_1 + z_2 + z_4 \\ \text{and } L'' &= g + (z_0 + z_2 + z_3 + z_5)(z_0 + z_1 + z_2 + z_4 + z_5) = z_5 \end{aligned}$$

The Boolean element g takes the value 0 on 31 of the 63 nonzero points. On these 31 nonzero 0-points for g, the linear Boolean elements

$$L = z_0 + z_2 + z_3, \quad L' = z_0 + z_1 + z_2 + z_4 \quad \text{and} \quad L'' = z_5$$

each take the value 0 for 23 of these 31 0-points of g, that is to say with approximate probability $\frac{3}{4}$ as stated in Lemma 5. Furthermore, all other nontrivial linear Boolean elements take the value 0 for 15 of the 31 nonzero 0-points of g, apart from $L + L' + L'' = z_1 + z_3 + z_4 + z_5$ which takes the value 0 for 7 of the 31 nonzero 0-points of g. □

4.5 Boolean Ring Equation Solving as an LPN Problem

We now discuss how to use the Boolean image of a homogeneous polynomial equation system to construct a *Learning Parity with Noise* or LPN problem [7], a standard and fundamental cryptographic problem.

We consider a system of quadratic Boolean elements $g_1 = 0, \ldots, g_{m'} = 0$, where $g_1, \ldots, g_{m'}$ satisfy the $\mathcal{R}_2$-Criterion and are obtained by considering the Boolean image under Γ of some original homogeneous polynomial equation system $f_1 = \ldots = f_m = 0$ with a single (for simplicity) nonzero solution z^*. We note that any such derived linear Boolean elements can be used to make a substitution to simplify the original equation system, and so we only consider quadratic Boolean elements without loss of generality. The ideas of Section 4.3 show that we can potentially find m_0 linear Boolean elements $L_1, \ldots, L_{m_0}$ each taking the value 0 at z^* with approximate probability $\frac{3}{4}$. If we regard $L_1, \ldots, L_{m_0}$ as (column) vectors of coefficients, then we can write such probabilistic linear expressions as

$$0 = L_j^T z^* + \epsilon_j, \text{ where } \mathbf{P}(\epsilon_j = 0) \approx \tfrac{3}{4} \text{ and } \mathbf{P}(\epsilon_j = 1) \approx \tfrac{1}{4} \quad [j = 1, \ldots, m_0],$$

where $\epsilon_1, \ldots, \epsilon_{m_0}$ are independent and identically distributed random variables on GF(2). For the $m_0 \times n$ matrix $C = \left(K_1^T | \ldots | K_{m_0}^T\right)^T$ over GF(2) and error vector $\epsilon = (\epsilon_1, \ldots, \epsilon_{m_0})^T$, we can write the probabilistic linear Boolean expressions in matrix form as the *statistical linear model*

$$0 = Cz^* + \epsilon.$$

The problem of determining z^* in the above probabilistic expression is an instance of the `LPN` problem, and so we can potentially use the `BKW` algorithm to address this `LPN` problem [7]. The `BKW` algorithm is essentially a form of Gaussian elimination in which the number of row additions is minimised is order to constrain the growth of the error rate, and we examine its use in this case. Without loss of generality, we assume that $n = ab$ and that $m_0 > 2^b$ and partition the matrix C as

$$C = (C_1 | \, C_2 | \, \ldots \, | C_a) \, ,$$

that is to say into $(m_0 \times b)$ submatrices $C_1, \ldots, C_a$. We then find distinct pairs j', j'' such the corresponding rows $C_{1j'}^T$ and $C_{1j''}^T$ of rows of C_1 are identical, so $C_{1j'}^T = C_{1j''}^T$. We can therefore construct a row vector

$$C_{j'}^T + C_{j''}^T = \left(0 \, \middle| \, C_{2j'}^T + C_{2j''}^T \middle| \, \ldots \, \middle| \, C_{aj'}^T + C_{aj''}^T\right)$$

in which the first b components are all 0. By constructing m_1 such vectors, we can obtain an $m_1 \times n$ matrix

$$C^{(1)} = \left(0 \, \middle| \, C_2^{(1)} \middle| \ldots \middle| \, C_a^{(1)}\right),$$

in which the left-most b columns are 0, giving the statistical linear model

$$0 = C^{(1)} z^* + \epsilon^{(1)},$$

in which a component $\epsilon_j^{(1)}$ of the new error $\epsilon^{(1)}$ is the sum of two of the components of the previous error vector ϵ. Thus $\mathbf{P}(\epsilon_j^{(1)} = 0) \approx \frac{3}{4}\frac{3}{4} + \frac{1}{4}\frac{1}{4} = \frac{5}{8}$ and $\mathbf{P}(\epsilon_j^{(1)} = 1) \approx \frac{3}{8}$. By iterating the process, we can obtain an $m_{a-1} \times n$ matrix

$$C^{(a-1)} = \left(0 \, \middle| \ldots \middle| \, 0 \, \middle| \, C_a^{(a-1)} \, \right),$$

in which the left $(a-1)b$ columns are 0, giving the statistical linear model

$$0 = C^{(a-1)} z^* + \epsilon^{(a-1)} = C_a^{(a-1)} z_{(a)}^* + \epsilon^{(a-1)}$$

where $z_{(a)}^* = (z_{(a-1)b+1}^*, \ldots, z_{ab}^*)^T$ is a vector of the final b components of z^*. In this case, a component $\epsilon_j^{(a-1)}$ of this new error is the sum of 2^{a-1} components of the original error, so these components are usually pairwise independent with $\mathbf{P}\left(\epsilon_j^{(a-1)} = 0\right) \approx \frac{1}{2}(1 + 2^{-a})$ and $\mathbf{P}\left(\epsilon_j^{(a-1)} = 1\right) \approx \frac{1}{2}(1 - 2^{-a})$. If this distribution is sufficently non-uniform, then we can accurately determine $z_{(a)}^*$ and so reduce the problem to an $(n-b)$-dimensional problem and so on.

4.6 Required Degree for the EGHAM2 Process to Succeed

A major determination of the complexity of the EGHAM2 process is the degree D of the underlying polynomial system, and in particular the minimal degree D required for the EGHAM2 process to complete without generating new polynomials of higher degree.

The degree D to which the original equation systems need to be extended for the comparable XL or Gröbner Basis algorithms to give a solution is considered by [1, 3]. Loosely speaking, these papers taken together argue that for most sets of m homogeneous polynomials of degree d, the minimal value of D' for which the coefficient of $y^{D'}$ is negative in the expansions of the expressions

$$\frac{(1+y)^{n'}}{(1+y^d)^m} \text{ for Gröbner Basis } \mathtt{F}_5 \text{ and } \frac{(1+y)^{n'}}{(1+y^d)^m} - \frac{1+y}{1-y} \text{ for } \mathtt{XL},$$

where $n' = n + 1$ is the number of variables, gives the required degree D. The EGHAM2 process though requires a degree for which a Rank-2 Product Polynomial can be found. We observe that the set of Rank-2 Product Polynomials contains subspaces of $\mathcal{B}$ of dimension $n+1$, for example $\langle x_0 x_i | i = 0, \ldots, n \rangle$. This suggests that the degree D required for the EGHAM2 process would in general be bounded by the degree D required for the XL algorithm for the same system.

4.7 An Example of the EGHAM2 Process

We illustrate the EGHAM2 process in Example 8, where we discuss a multivariate quadratic system over GF(2). Whilst this system is relatively small (it could easily be solved by exhaustive search), it does demonstrate the advantages of the EGHAM2 process in comparison with an XL or Gröbner basis approach for such a multivariate GF(2)-system. Furthermore, we discuss we can use the Boolean image of the system to generate an LPN instance and how to use the BKW algorithm to solve this LPN instance.

Example 8. We consider as example with $m = 63$ randomly generated homogeneous quadratic equations in $n' = 20$ variables over GF(2). There are 210 homogeneous monomials of degree 2, so each such polynomial consists of about 105 homogeneous quadratic terms. In this case, the "XL-polynomial" of Section 4.6 expands as

$$\frac{(1+y)^{20}}{(1+y^2)^{63}} - \frac{1+y}{1-y} = 19y + 145y^2 - 120y^3 + \ldots,$$

so indicating that it should be possible to obtain cancellation with by generating cubic homogeneous polynomials from these 63 polynomial equations in 20 variables. However, there is generally no bivariate polynomial in the span of the resulting 1323 (1260 cubic and 63 quadratic) generated cubic polynomials, so an XL or a Gröbner basis approach would typically require the generation of quartic polynomials.

The EGHAM2 process by contrast can solve this quadratic equation system with 63 quadratic polynomial equation systems in 20 variables whilst only generating cubic polynomials and not using any quartic polynomials. In a typical instance, the 1323 generated cubic polynomials contained 183 polynomials satisfying the $\mathcal{Q}$-Criterion, that is to say polynomials whose image under the Boolean mapping Γ is a quadratic Boolean element.

Given such cubic polynomials with quadratic Boolean images, the EGHAM2 approach uses the Kernel Method to find ∂-matrices of rank 2 in the span of the ∂-matrices arising from the 183 polynomials satisfying the $\mathcal{Q}$-Criterion. This approach proceeds by determining the (left) kernel of the $m \times l(n+1)$ matrix $\left(\Lambda^{(v_1)}\middle|\ldots\middle|\Lambda^{(v_l)}\right)$, and the usual dimension of this kernel is given for various values of l below.

l	1	2	3	4	5	6	7	8	9	10	11	12	13
Kernel Dimension	164	146	129	113	98	84	71	59	48	38	29	21	14

We make in passing the following observation for the dimension of this kernel as l increases. We originally considered 183 polynomials, we can technically regard the kernel dimension for $l = 0$ corresponding to a "183×0" matrix with kernel of dimension 183. The kernel dimension for $l = 1$ is 164, which is $n = 19$ less than 183. The kernel dimension for $l = 2$ is 146, which is 18 less than 146 and so on. For this example, we use the above values to choose $l = 13$ generally giving rise to a 14-dimensional kernel for the 183×260 matrix $\left(\Lambda^{(v_1)}\middle|\ldots\middle|\Lambda^{(v_{13})}\right)$.

We used 500 iterations of Kernel Method with $l = 13$, that is to say we generated 500 matrices of the above form $\left(\Lambda^{(v_1)}\middle|\ldots\middle|\Lambda^{(v_{13})}\right)$, to find ∂-matrices of rank 2, that is to say quadratic Boolean elements satisfying the $\mathcal{R}_2$-Criterion. No linear Boolean elements were found, and each quadratic Boolean element found satisfying the $\mathcal{R}_2$-Criterion can in practice be used to give three probabilistic linear expressions for the solution. These iterations of the Kernel method gave $m_0 = 1905$ linear Boolean expressions each taking the value 0 (with the true z^*) with probability approximately $\frac{3}{4}$. Thus we can obtain the statistical linear model $0 = Cz^* + \epsilon$ over GF(2) with $m_0 \times n'$ or 1905×20 matrix C and $\mathbf{P}(\epsilon_i = 0) \approx \frac{3}{4}$, so giving an instance of the LPN Problem.

This instance of the LPN problem can be solved by implementing the BKW algorithm by taking $a = 2$ and $b = 10$, that is to say by dividing $C = (C_1|C_2)$ into two $m_0 \times b$ or 1905×10 submatrices C_1 and C_2. As $m_0 > 2^b$, the BKW algorithm reduces the left half of the columns to 0 to give a $m_1 \times b$ matrix $C_{(2)}^{(a-1)}$, where $m_1 = 1037$ in this case. Thus the BKW algorithm gives a 10-dimensional statistical linear model $0 = C_2^{(1)} z_{(2)}^* + \epsilon^{(1)}$ over GF(2), where $z_{(2)}^*$ is the "right half" of solution z^* and $\mathbf{P}\left(\epsilon^{(1)} = 0\right) \approx \frac{5}{8}$.

The true value of $z_{(a)}^*$ can then be identified by evaluating the $2^b - 1 = 2^{10} - 1$ counters

$$S_{z'} = m_1 - \text{Wt}\left(C_{(2)}^{(1)} z'\right) = 1037 - \text{Wt}\left(C_{(2)}^{(1)} z'\right)$$

for $z' \neq 0$ giving the number of 0-components of the vector $C_2^{(1)} z$ of dimension $m_1 = 1037$. The distribution of these counts when $z' = z_{(a)}^*$ takes the correct

value and $z' \neq z^*_{(a)}$ takes an incorrect value are given by

$$\begin{aligned} S_z &\sim \mathrm{Bin}(1037, \tfrac{5}{8}) \approx \mathrm{N}(648.1, 15.6^2) \quad [z = z^*_{(a)}] \\ \text{and } S_z &\sim \mathrm{Bin}(1037, \tfrac{1}{2}) \approx \mathrm{N}(518.5, 16.1^2) \quad [z \neq z^*_{(a)}]. \end{aligned}$$

In essence, we can identify the true value of $z^*_{(a)}$ if a realisation of $\mathrm{N}(648.1, 15.6^2)$ distribution exceeds the maximum of 1023 realisations of a $\mathrm{N}(518.5, 16.1^2)$ distribution. More generally, an accurate probability for identifying the partial true solution can be determined by techniques using order statistics, as discussed in a cryptographic context by [9]. In this case, the partial true solution immediately identifies itself with an $S_{z'}$-count of 652 compared with the next highest $S_{z'}$-count of 567. Making the appropriate substitutions then gives a polynomial equation system with 63 quadratic polynomial equations in 10 variables. which is a fully linearised system that can be solved directly. Thus the system of 63 quadratic polynomial equations in 20 variables can be solved by the `EGHAM2` process using only cubic monomials and basic linear algebra. □

5 Conclusions

We have outlined a new method, the `EGHAM2` process, specifically designed for analysing polynomial systems over GF(2) that have a small number of solutions. This method is expected to be more efficient that the comparable `XL` or Gröbner Basis methods for the following reasons.

- The `EGHAM2` process is geometrically invariant, whereas the comparable `XL` or Gröbner basis algorithms are in general not geometrically invariant.
- The degree D required by the `EGHAM2` process for the extended polynomial system should be bounded by the degree required by the `XL` or Gröbner Basis algorithms.
- The processing required by the `EGHAM2` algorithm should be more straightforward as it is focussed on a much smaller quadratic system. The `EGHAM2` algorithm also avoids the possible complexities involved in testing trial roots of high degree polynomials generated by `XL` or Gröbner Basis algorithms.

Furthermore, the `EGHAM2` process establishes a direct natural connection between solving a multivariate polynomial equation system over GF(2), an instance of an `MQ` problem, and solving an instance of an `LPN` problem.

Acknowledgements

We would like to thank the anonymous referees for their helpful comments.

References

1. G. Ars, J-C. Faugère, H. Imai, M. Kawazoe, and M. Sugita. Comparison Between XL and Gröbner Basis Algorithms. In P-J. Lee, editor, *Advances in Cryptology - ASIACRYPT 2004*, volume 3329 of *LNCS*, pages 338–353. Springer, 2004.

2. M. Atiyah and I. MacDonald. *Introduction to Commutative Algebra*. Westview Press, 1994.
3. M. Bardet, J.C. Faugère, and B. Salvy. Complexity of Gröbner Basis Computation for Semi-Regular Overdetermined sequences over GF(2) with Solutions in GF(2). Technical report, INRIA Research Report 5049, available at `http://www-polsys.lip6.fr/~jcf/Papers/RR-5049.pdf`, 2003.
4. M. Bardet, J.C. Faugère, and B. Salvy. On the Complexity of Gröbner Basis Computation of Remi-Regular Overdetermined Algebraic Equations. In *International Conference on Polynomial System Solving - ICPSS*, pages 71 –75, 2004.
5. L. Bettale, J-C. Faugère, and L. Peret. Cryptanalysis of HFE, Multi-HFE and Variants for odd and even Characteristic. *Designs, Codes and Cryptography*, 69:1–52, 2013.
6. O. Billet and J. Ding. Overview of Cryptanalysis Techniques in Multivariate Public Key Cryptography. In *Gröbener Bases, Coding and Cryptography*. Springer, 2009.
7. A. Blum, A. Kalai, and H. Wasserman. Noise-Tolerant Learning, the Parity Problem, and the Statistical Query Model. *Journal of the ACM*, 50:506–519, 2003.
8. C. Bouillaguet, P-A. Fouque, and G. Macario-Rat. Practical Key-Recovery for all possible Parameters of SFLASH. In D.H. Lee and X. Wang, editors, *Asiacrypt 2011*, volume 7073 of *LNCS*, pages 667–685. Springer, 2011.
9. R. Bricout, S. Murphy, K Paterson, and T. van der Merwe. Analysing and Exploiting the Mantin Biases in RC4. *Designs, Codes and Cryptography*, 84:743–770, 2018.
10. B. Buchberger. *Ein Algorithmus zum Auffinden der Basiselemente des Restklassenringes nach einem nulldimensionalen Polynomideal (An Algorithm for Finding the Basis Elements in the Residue Class Ring Modulo a Zero Dimensional Polynomial Ideal)*. PhD thesis, Mathematical Institute, University of Innsbruck, Austria, 1965. English translation in J. of Symbolic Computation, Special Issue on Logic, Mathematics, and Computer Science: Interactions. Vol. 41, Number 3-4, Pages 475–511, 2006.
11. J.A. Buchmann, J. Ding, M.S.E Mohamed, and W. Mohamed. MutantXL: Solving Multivariate Polynomial Equations for Cryptanalysis. In H. Handschuh, S. Lucks, B. Preneel, and P. Rogaway, editors, *Symmetric Cryptography*, volume 09031 of *Dagstuhl Seminar Proceedings*, 2009.
12. N. Courtois. Higher Order Correlation Attacks, XL Algorithm and Cryptanalysis of Toyocrypt. In P-J. Lee and C. Lim, editors, *Information Security and Cryptology - ICISC 2002*, LNCS, pages 182–199, 2002.
13. N. Courtois and L. Goubin. Cryptanalysis of the TTM Cryptosystem. In T. Okamoto, editor, *Asiacrypt 2000*, volume 1976, pages 44–57, 2000.
14. N. Courtois, L. Goubin, and J. Patarin. SFLASHv3, a fast asymmetric signature scheme. *IACR Cryptology ePrint Archive*, 2003:211, 2003.
15. N. Courtois, A. Klimov, J. Patarin, and A. Shamir. Efficient Algorithms for Solving Overdefined Systems of Multivariate Polynomial Equations. In B. Preneel, editor, *Advances in Cryptology - EUROCRYPT 2000*, volume 1807 of *LNCS*, pages 392–407. Springer, 2000.
16. N. Courtois and J. Patarin. About the XL Algorithm over GF(2). In M. Joye, editor, *Topics in Cryptology - CT-RSA 2003*, LNCS, pages 141–157, 2003.
17. N. Courtois and J. Pieprzyk. Cryptanalysis of Block Ciphers with Overdefined Systems of Equations. In Y. Zheng, editor, *Advances in Cryptology - ASIACRYPT 2002*, LNCS, pages 267–287, 2002.

18. C. Diem. TheXL-Algorithm and a Conjecture from Commutative Algebra. In P.J. Lee, editor, *Advances in Cryptology - ASIACRYPT 2004*, volume 3329 of *LNCS*, pages 323–337. Springer, 2004.
19. J. Ding and B-Y. Yang. *Post-Quantum Cryptography*, chapter Multivariate Public Key Cryptography, pages 193–241. Springer, 2009.
20. J-C. Faugère. A New Efficient Algorithm for Computing GröBner Bases without Reduction to Zero (F5). In *Proceedings of the 2002 International Symposium on Symbolic and Algebraic Computation*, ISSAC'02, pages 75–83. ACM, 2002.
21. J-C. Faugère, F. Levy dit Vehel, and L. Perret. Cryptanalysis of MinRank. In D. Wagner, editor, *Advances in Cryptology - CRYPTO '08*, volume 5157 of *LNCS*, pages 280–296. Springer, 2008.
22. J-C. Faugère, D. Gligoroski, L. Perret, S. Samardjiska, and E. Thomae. A Polynomial-Time Key-Recovery Attack on MQQ Cryptosystems. In J. Katz, editor, *Public-Key Cryptography - PKC 2015*, LNCS, pages 150–174. Springer, 2015.
23. J-C. Faugère and A. Joux. Algebraic Cryptanalysis of Hidden Field Equation (HFE) Cryptosystems Using Gröbner Bases. In D. Boneh, editor, *Advances in Cryptology - CRYPTO 2003*, LNCS, pages 33–60, 2003.
24. A. Fraenkel and Y. Yesha. Complexity of Solving Algebraic Equations. *Information Processing Letters*, 10:178–179, 1980.
25. A. Kipnis and A. Shamir. Cryptanalysis of the Oil and Vinegar Signature Scheme. In H. Krawcyzk, editor, *Advances in Cryptology - CRYPTO '98*, volume 1462 of *LNCS*, pages 257–266. Springer, 1998.
26. A. Kipnis and A. Shamir. Cryptanalysis of the HFE Public Key Cryptosystem by Relinearization. In M. Wiener, editor, *Advances in Cryptology - CRYPTO '99*, volume 1666 of *LNCS*, pages 19–30. Springer, 1999.
27. T. Matsumoto and H. Imai. Public Quadratic Polynominal-Tuples for Efficient Signature-Verification and Message-Encryption. In C. Günther, editor, *Advances in Cryptology - EUROCRYPT '88*, volume 330 of *LNCS*, pages 419–453. Springer, 1988.
28. S. Murphy and M.B. Paterson. A Geometric View of Cryptographic Equation Solving. *Journal of Mathematical Cryptology*, 2(1):63–107, 2008.
29. S. Murphy and M.B. Paterson. Geometric Ideas for Cryptographic Equation Solving in Even Characteristic. In M. Parker, editor, *Cryptography and Coding, 12th IMA International Conference 2009*, LNCS, pages 202–221. Springer, 2009.
30. National Institute of Science and Technology (NIST). Post-Quantum Cryptographic Standardization Process. Technical report, available at `https://csrc.nist.gov/projects/post-quantum-cryptography`, 2017-20.
31. J. Patarin. Hidden Fields Equations (HFE) and Isomorphisms of Polynomials (IP): Two New Families of Asymmetric Algorithms. In U. Maurer, editor, *Advances in Cryptology - EUROCRYPT '96*, volume 1070 of *LNCS*, pages 33–48, 1996.
32. J. Patarin. Cryptanalysis of the Matsumoto and Imai Public Key Scheme of Eurocrypt'98. *Designs, Codes and Cryptography*, 20:175–209, 2000.
33. M. Sugita, M. Kawazoe, L. Perret, and H. Imai. Algebraic Cryptanalysis of 58-Round SHA-1. In A. Biryukov, editor, *Fast Software Encryption FSE 2007*, volume 4593 of *LNCS*, pages 349–365. Springer, 2007.
34. B-Y. Yang and J-M. Chen. All in the XL Family: Theory and Practice. In C. Park and S. Chee, editors, *Information Security and Cryptology - ICISC 2004*, volume 3506 of *LNCS*, pages 67–86. Springer, 2004.