Cryptanalysis of HFE with Internal Perturbation

Vivien Dubois, Louis Granboulan, and Jacques Stern*

École normale supérieure
DI, 45 rue d'Ulm, 75230 Paris cedex 05, France
{vivien.dubois,louis.granboulan,jacques.stern}@ens.fr

Abstract. Multivariate Cryptography has been an active line of research for almost twenty years. While most multivariate cryptosystems have been under attack, variations of the basic schemes came up as potential repairs. In this paper, we study the Internal Perturbation variation of HFE recently proposed by Ding and Schmidt. Although several results indicate that HFE is vulnerable against algebraic attacks for moderate size parameters, Ding and Schmidt claim that the cryptosystem with internal perturbation should be immune against them. However in this paper, we apply the recently discovered method of differential analysis to the Internal Perturbation of HFE and we find a subtle property which allows to disclose the kernel of the perturbation. Once this has been achieved, the public key can be inverted by attacking the underlying HFE provided the parameters were taken low enough to make the perturbed scheme of competitive performance.

Keywords: multivariate cryptography, HFE, internal perturbation, differential cryptanalysis, binary vector spaces.

1 Introduction

Multivariate Cryptography has been an active line of research for almost twenty years. Initiated independently in the early 80's by Matsumoto-Imai and Fell-Diffie [11,7], the field was revived by the work of Patarin and Shamir [14,17,15]. The interest for multivariate primitives can be explained in several ways. First, these schemes are not related to factorization or discrete logarithm problems. They rely on the intractability of solving systems of multivariate quadratic equations over a finite field. This problem is proved NP-hard [12] and moreover no quantum polynomial algorithm has been found to solve it. Next, these schemes benefit from several nice properties such as providing very short or very fast signatures, as well as a very particular flexibility: from all basic trapdoors can be derived a number of generic variations. These variations are often considered to thwart structural attacks against the original cryptosystems.

Today most basic trapdoors have been under attack. Among the most promising, HFE was introduced by Patarin as a repair of the Matsumoto-Imai cryptosystem [15]. The scheme was quickly subject to a cryptanalytic attack by Kipnis and Shamir [9], and further attacked by Courtois [1], but the first successful

* This work is supported in part by the French government through X-Crypt, in part by the European Commission through ECRYPT

cryptanalysis of HFE was only provided by Faugère and Joux, eight years after its invention [6]. The latter attack made use of a general Gröbner bases algorithm and its success can only be explained by some inherent algebraic properties allowing a peculiarly fast computation of the algorithm. These algebraic properties were recently mathematically explained by Granboulan-Joux-Stern [10] and a rather clear picture of how to choose parameters to withstand attacks is emerging.

On the other hand, very few studies are dedicated to the security of variations, and the respective effects of the many variations remain unclear in terms of security. However variations are powerful and can have a crucial impact on security: as an example, the SFlash signature algorithm chosen by the NESSIE European consortium is a variation of the broken Matsumoto-Imai cryptosystem [13]. Also, most attacks against the basic cryptosystems do not extend to variations. The gain on security brought by variations has to be understood to determine whether they result in secure schemes.

Our results In this paper, we consider a variation of HFE called the Internally Perturbed HFE. This variation was recently proposed by Ding and Schmidt [4]. It was designed to counter Kipnis-Shamir's attack, and is expected to withstand Gröbner bases attack as well. A simpler internal variation had been previously proposed based on the Matsumoto-Imai cryptosystem [2] and had already been asserted to provide immunity against algebraic attacks [3]. Unfortunately, the Matsumoto-Imai cryptosystem has a very specific structure and the internal perturbation could actually be removed using the recently introduced differential technique [8]. In this work, we consider the enhanced internal perturbation variation as applied to HFE and defined in [4]. We show that the original internal perturbation variation applied to HFE still suffers from the drawback exhibited in [8], while the enhanced version has indeed a much subtler differential visibility. However, a differential bias can still be captured and exploited to disclose the kernel of the perturbation. Once this has been achieved, the public key can be inverted by attacking the underlying HFE provided the parameters were taken low enough to make the perturbed scheme of competitive performance. Precise complexity estimates for the attack are provided.

Organization of the paper In section 2, we recall the construction of HFE and its Internal Perturbation variation. Next, in section 3, we recall the basics of differential analysis for multivariate schemes and its application to the internally Perturbed Matsumoto-Imai. In section 4, we analyze the differential of the Internally Perturbed HFE and we exhibit a provable distinguisher of elements cancelling the perturbation. In section 5, we turn this distinguisher into an algorithm to find the kernel of the perturbation. In section 6, we show that the public key can be easily inverted once this kernel is known. The method being quite technical in character, all proofs could not be included; the full paper is available from the authors.

2 The Internally Perturbed HFE cryptosystem

2.1 Notations

We denote by $\mathbb{F}_2$ the finite field with two elements and by $\mathbb{F}_{2^n}$ the degree n extension field of $\mathbb{F}_2$. $\mathbb{F}_{2^n}$ is an $\mathbb{F}_2$-vector space of dimension n isomorphic to $\mathbb{F}_2^n$. The squaring operation $x \mapsto x^2$ is $\mathbb{F}_2$-linear (or additive) in $\mathbb{F}_{2^n}$. As a consequence, sums of monomials of the form ax^{2^i} where a is an element of $\mathbb{F}_{2^n}$ and i is an integer in $[0, n-1]$, are the $\mathbb{F}_2$-linear maps over $\mathbb{F}_{2^n}$. Polynomials of this type will be therefore called $\mathbb{F}_2$-*linear polynomials*. Given an $\mathbb{F}_2$-linear polynomial, the set of its cancelling elements is a linear subspace of $\mathbb{F}_{2^n}$ that will be referred to as its *kernel*. $\mathbb{F}_2$-linear polynomials are isomorphic to (multivariate) linear maps of $\mathbb{F}_2^n$ by an extension of the isomorphism between $\mathbb{F}_{2^n}$ and $\mathbb{F}_2^n$. Similarly, sums of monomials of the form $ax^{2^i+2^j}$ where a is an element of $\mathbb{F}_{2^n}$ and i, j are integers in $[0, n-1]$, will be called $\mathbb{F}_2$-*quadratic polynomials*. $\mathbb{F}_2$-quadratic polynomials translate through the isomorphism between $\mathbb{F}_{2^n}$ and $\mathbb{F}_2^n$ into quadratic maps of $\mathbb{F}_2^n$, defined by n polynomials of degree 2 in n variables.

2.2 The Original HFE Setting

Informally speaking, the generic construction of multivariate schemes consists in disguising an easily solvable system of multivariate quadratic equations as random, by a secret transformation. In most schemes, the secret transformation is the composition by two randomly chosen invertible affine maps S, T; one is applied on the variables and the other one on the equations. The way to generate an easily solvable quadratic system P defines each scheme. The public key $\boldsymbol{P}$ is given by:

$$\boldsymbol{P} = T \circ P \circ S$$

An encrypted message $\boldsymbol{P}(a)$ is decrypted by solving the quadratic system $\boldsymbol{P}(x) = \boldsymbol{P}(a)$. Solving this system is intractable except for the legitimate user which can invert T and S and solve the easy internal system. In Matsumoto-Imai and HFE, the easily solvable system P exploits the isomorphism between $\mathbb{F}_{2^n}$ and $\mathbb{F}_2^n$. In Matsumoto-Imai, the internal function P is the multivariate expression of an $\mathbb{F}_2$-quadratic monomial $x^{2^i+2^j}$, where i, j are suitably chosen so that it is invertible. In HFE, the internal polynomial is the multivariate expression of an $\mathbb{F}_2$-quadratic polynomial which has low degree to allow decryption by a root-finding algorithm.

Different cryptanalytic approaches [9,6,1] made clear that the low degree of the internal polynomial in HFE makes the system vulnerable to algebraic attacks. In particular, Faugère and Joux demonstrated that systems of quadratic equations coming from HFE public keys allow much easier Gröbner basis computations than random systems of the same size [6] - the first challenge of HFE of parameters $n = 80$ and degree 96 was broken in a hundred hours. Now the question is : how to enhance the security of HFE ?

2.3 The Internally Perturbed HFE

To withstand low degree attacks, the internal polynomial should be modified so that it no more has low degree while still allowing decryption. An interesting idea to realize this, was presented by Ding and Schmidt [4] and is known as the Internally Perturbed HFE. The suggested modification consists in "noising" the low degree internal polynomial by a few terms of high degree which can only be removed by the legitimate user. We next recall this scheme in detail.

For a given degree parameter D, the user chooses a bivariate polynomial $\ddot{P}(x, y)$ as the sum of three basic components:

- a univariate $\mathbb{F}_2$-quadratic polynomial $P(x)$ in variable x of low degree under 2^{D+1}, that will be called *the HFE-part of* $\ddot{P}$.
- a bivariate $\mathbb{F}_2$-bilinear polynomial $M(x, y)$ in variables x, y of low degree 2^D in x, that will be called the *the mixing part of* $\ddot{P}$.
- a univariate $\mathbb{F}_2$-quadratic polynomial $\overline{P}(y)$ in variable y, that will called *the pure perturbation part of* $\ddot{P}$.

In addition, the user randomly selects an $\mathbb{F}_2$-linear polynomial $Z(x)$ of low rank r. The $\mathbb{F}_2$-quadratic polynomial $\tilde{P}(x) = \ddot{P}(x, Z(x))$ has very high degree in general, nevertheless its roots can be found indirectly: the image of Z, that we note $\mathrm{Im}(Z)$, has only 2^r elements and for any b of them, one can find the roots of $\ddot{P}(x, b)$ since it has small degree. $\tilde{P}(x)$ consists in the internal polynomial in the Internally Perturbed HFE, and the public key is $\tilde{\boldsymbol{P}} = T \circ \tilde{P} \circ S$, as in HFE. One can observe that the decryption process is 2^r times slower than for an HFE of the same degree parameter. The prescribed parameters are $n = 89, D = 3, r = 2$ [4]. It can be noticed that in our definition of the internal polynomial $\tilde{P}$, all linear and constant terms of the definition of [4] were omitted. Indeed in the sequel, we will only be interested in the differential of $\tilde{P}$, and as we will see, linear and constant terms disappear when taking the differential.

3 Internal Perturbation and Differential Analysis

We let $\boldsymbol{Z}$ to be the composition of Z with the linear part of S. As a basic observation, an Internally Perturbed HFE public key is just an HFE public key on any affine subspace parallel to the kernel of $\boldsymbol{Z}$. Indeed, this is required by the decryption process: for any element b, $\tilde{P}(x)$ coincides with the small degree polynomial $\ddot{P}(x, b)$ over the affine subspace $b + \ker Z$. Therefore, if we could discover the kernel of $\boldsymbol{Z}$, we could invert the public key by attacking the underlying HFEs with Gröbner bases, as shown by Faugère and Joux [6]. Hence, the Internally Perturbed scheme would be broken by the ability to recover the kernel of the perturbation.

Differential Analysis is a generic tool of analysis of multivariate schemes which can allow learning information about the hidden structure. It was in particular used to discover the kernel of the perturbation of a former internally perturbed scheme, the Perturbed Matsumoto-Imai cryptosystem. We next recall the basics of differential analysis for multivariate schemes and its application to the Perturbed Matsumoto-Imai.

3.1 Basic Properties of the Differential of a Quadratic Function

For any quadratic function P and any element a, the difference $P(x+a) - P(x)$ is an affine function in x of constant term $P(a) - P(0)$. Its linear part is called the differential of P at a and will be denoted DP_a in the sequel.

$$DP_a(x) = P(a+x) - P(x) - P(a) + P(0)$$

In multivariate schemes, we have two quadratic functions $\boldsymbol{P}$ and P which are related by two bijective affine transforms S and T following $\boldsymbol{P} = T \circ P \circ S$. Denoting $\underline{S}$ and $\underline{T}$ the linear parts of S and T, the differential of $\boldsymbol{P}$ and P are related the following way:

$$D\boldsymbol{P}_a = \underline{T} \circ DP_{\underline{S}(a)} \circ \underline{S}$$

Therefore, $\underline{S}$ and $\underline{T}$ being invertible, the distribution of the kernel-dimension of the differential for a random a is the same for the public key as for the internal function. This was first noticed in [8] to attack the Perturbed Matsumoto-Imai.

3.2 Application to the Perturbed Matsumoto-Imai

The Matsumoto-Imai scheme uses an internal polynomial P of the form $x^{2^i+2^j}$. Ding proposed an internal perturbation with *no mixing part* (i.e. $M(x,y) = 0$) [2]. Considering the differential of $\tilde{P}$ at a,

$$D\tilde{P}_a(x) = DP_a(x) + D\bar{P}_{Z(a)}(Z(x)) \tag{1}$$

it was observed in [8] that the differential at points in the kernel of Z is exactly the differential of the original Matsumoto-Imai function at these points. Besides, the differential of the Matsumoto-Imai internal function $x^{2^i+2^j}$ has kernel-dimension $gcd(n, i-j)$ at any non-zero point. On the other side, when taken at a point which is not in the kernel of Z, the perturbation part interferes and may cause the differential to have a larger of smaller kernel. This provides an easy criteria to detect elements which are not in the kernel of Z, and with sufficiently many such points, the kernel can be recovered.

As a remark, observe that the internal perturbation without mixing terms applied on HFE yields the same drawback. Again the differential of $\tilde{P}$ at a point of the kernel of Z is the differential of the HFE internal polynomial. The differential of an HFE internal polynomial of degree under 2^{D+1} has degree at most 2^D, and therefore, as a linear map, its kernel has dimension at most D [5]. On the other side, when the perturbation interferes, the differential may have a larger kernel.

4 A Differential Bias of the Internally Perturbed HFE

In this section, we prove the spinal cord of our attack: whether the perturbation vanishes or not yields a differential bias. First, we characterize the form of the differential in both cases, in terms of sums of linear maps of two kinds. Second, we compute the distribution of the kernel-dimensions in both cases, using combinatorics in binary vector spaces. Third, we define a distinguisher of kernel elements whose advantage can be exactly computed for a random secret key.

4.1 Differential Structure of the Perturbed Internal Polynomial

>From now on, the kernel of Z will be denoted $\mathcal{K}$. Depending on the membership of a to $\mathcal{K}$, the differential at a is:

$$\begin{array}{llllll} a \notin \mathcal{K}, & D\tilde{P}_a(x) = DP_a(x) & +M(x, Z(a)) & +M(a, Z(x)) & +D\bar{P}_{Z(a)}(Z(x)) \\ a \in \mathcal{K}, & D\tilde{P}_a(x) = DP_a(x) & & +M(a, Z(x)) & \end{array}$$

As we can see, the *differential of the perturbed internal at points where the perturbation vanishes is not the differential of the non-perturbed internal* as it was for PMI. In particular, the kernel-dimension of the differential will be more than D for some elements in $\mathcal{K}$ while this could never happen with a PMI-like perturbation. In fact, we will next show that the differential reaches the same kernel dimensions in both cases. Therefore, it will not be possible to use the "cut-off" based strategy as for PMI to detect the effectiveness of the perturbation. A more elaborate analysis of the differential is therefore required.

As a first step, we can observe that the structure of the differential is very similar in both cases. In both cases, this is the sum of an $\mathbb{F}_2$-linear polynomial of degree 2^D and a linear map of rank r which take the same value at a. What differs is: the common value is 0 when a is in $\mathcal{K}$ and non-zero when it is not. This actually captures the structure of this differential, as stated by the following theorem.

Theorem 1. *Let a be a non-zero element of $\mathbb{F}_2^n$. A random instance $(P, M, \bar{P}, Z)$ of Internally Perturbed HFE with parameters (D, r) has an internal polynomial denoted $\tilde{P}$. We denote by L_D a random $\mathbb{F}_2$-linear polynomial of degree 2^D and by l_r a random linear map of rank r. Then, for a proportion $1 - \epsilon_{n,r}$ of all instances $(P, M, \bar{P}, Z)$ of the cryptosystem, we have:*

$$Pr\left[\dim\ker D\tilde{P}_a = t \,|\, a \in \mathcal{K}\right] = Pr[\dim\ker(L_D + l_r) = t \,|\, L_D(a) = l_r(a) = 0]$$

and

$$Pr\left[\dim\ker D\tilde{P}_a = t \,|\, a \notin \mathcal{K}\right] = Pr[\dim\ker(L_D + l_r) = t \,|\, L_D(a) = l_r(a) \neq 0]$$

where $\epsilon_{n,r} = 2^{-(n-r)} + \mathcal{O}(2^{-2n})$.

A proof of the theorem can be found in the full paper available from the authors. It will be clear from the sequel that, for the suggested parameters, $\epsilon_{n,r}$ is negligible compared to the probabilities of interest. Accordingly, the kernel dimensions of the differential at points inside and outside $\mathcal{K}$ respectively follow the distributions of probability denoted π^+ and π^- defined by:

$$\pi^+(t) = \Pr_{(L_D,l_r)}\left[\dim\ker(L_D + l_r) = t \mid L_D(a) = l_r(a) = 0\right]$$
$$\pi^-(t) = \Pr_{(L_D,l_r)}\left[\dim\ker(L_D + l_r) = t \mid L_D(a) = l_r(a) \neq 0\right]$$

We next study both distributions in detail.

4.2 Distribution of the kernel-dimension of the differential depending on the position of the point

Distributions π^+ and π^- can be exactly computed using combinatorics in binary vector spaces, which are of independent interest. We will not recall these combinatorics here since they are not the subject of this paper, however all details are provided in Appendix A. We nevertheless describe the three steps that we follow to determine the distribution of the kernel-dimension of the sum of a random $\mathbb{F}_2$-linear polynomial of degree 2^D and a random linear map of rank r:

- first, we compute the distribution of the kernel-dimension of $\mathbb{F}_2$-linear polynomials of degree 2^D. The kernel-dimension of such polynomials is at most D, and the vanishing of one such polynomial over a subspace of dimension d with $d \leq D$ can be expressed in d independent linear constraints over the $D+1$ coefficients defining this $\mathbb{F}_2$-linear polynomial.
- fixing an $\mathbb{F}_2$-linear polynomial L of kernel-dimension d, we can compute the probability that a random subspace of dimension $n-r$ has intersection of dimension i with the kernel of L.
- fixing a subspace G of dimension $n-r$ which intersects the kernel of L with dimension i, we can enumerate the number of linear maps l of kernel G such that $\ker(L+l)$ has dimension t. Observe that in characteristic 2, $\ker(L+l)$ is the subspace where L and l are equal.

The overall probability for the dimension t requires to sum over all possible values of d and i; unfortunately, we could not find a closed formula (if any) for this probability. Nevertheless the sum itself is enough for all practical purposes.

Finding the laws π^+ and π^- consists in redoing the previous enumeration while taking into account the constraint at a. For any d and i, we can extract the correction factors coming from the constraint at a in either case. This leads to the following proposition.

Proposition 1. *Let $\pi_{d,r,i}(t)$ be the probability that the sum $L_D + l_r$ of a random $\mathbb{F}_2$-linear polynomial L_D of degree 2^D and kernel-dimension d and a random linear map l_r of rank r with kernels intersecting with dimension i, has kernel dimension t. Formally,*

$$\pi_{d,r,i}(t) = Pr_{(L_D,l_r)}\left[\dim\ker(L_D + l_r) = t\,;\begin{cases}\dim\ker L_D = d\\ \dim(\ker L_D \cap \ker l_r) = i\end{cases}\right]$$

For a prescribed non-zero element a, we denote

- $\pi^+_{d,r,i}(t)$ *for the probability of the same event knowing $L_D(a) = l_r(a) = 0$*
- $\pi^-_{d,r,i}(t)$ *for the probability of the same event knowing $L_D(a) = l_r(a) \neq 0$*

We have:

$$\begin{aligned}\pi^+_{d,r,i}(t) &= \quad 2^r \quad (2^i - 1)\ \pi_{d,r,i}(t)\ (1 + 2^{-(n-r)} + \mathcal{O}(2^{-2(n-r)}))\\ \pi^-_{d,r,i}(t) &= \tfrac{2^r}{2^r-1}\ (2^t - 2^i)\ \pi_{d,r,i}(t)\end{aligned}$$

on average over a.

Again, a proof can be found in the full paper. Neglecting terms of order $2^{-(n-r)}$, we obtain for π^+ and π^-:

$$\pi^+(t) = \quad 2^r \quad \sum_{d=0}^{D} \sum_{i=0}^{d} \left(2^i - 1\right) \pi_{d,r,i}(t)$$

$$\pi^-(t) = \tfrac{2^r}{2^r-1} \sum_{d=0}^{D} \sum_{i=0}^{d} \left(2^t - 2^i\right) \pi_{d,r,i}(t)$$

Though these probabilities are not provided under a closed form, they can be computed for any choice of the parameters. For example, for the suggested parameters $(n, D, r) = (89, 3, 2)$ their values are given in the table below:

dimension t	π^+	$\pi^-(t)$	$\text{sign}(\pi^+ - \pi^-)$
1	$\simeq 0.57764$	$\simeq 0.57756$	$+$
2	$\simeq 0.38495$	$\simeq 0.38507$	$-$
3	$\simeq 0.036718$	$\simeq 0.036662$	$+$
4	$\simeq 0.00069427$	$\simeq 0.00070045$	$-$
5	$\simeq 0.0000025431$	$\simeq 0.0000029064$	$-$

The kernel-dimension of the differential at some point a is now fully understood: it can follow two well determined distributions depending on the membership to $\mathcal{K}$ of a. We next compare these two distributions and show that the kernel-dimension of the differential at a yields some information about its membership or non-membership to $\mathcal{K}$.

4.3 Distinguishing Kernel Elements

Definition of our Distinguisher Let $\tilde{\boldsymbol{P}}$ be a public key associated to a given instance $(P, M, \bar{P}, Z)$ of the cryptosystem, and let $\boldsymbol{\mathcal{K}}$ be the subspace isomorphic to $\mathcal{K}$ through the linear masking. Our distinguisher is built on the differential bias exhibited in the preceding section. For a random non-zero element a, we compute the kernel dimension of the differential of $\tilde{\boldsymbol{P}}$ at a and obtain the dimension t. If for this dimension t we have $\pi^+(t) \geq \pi^-(t)$ then the hypothesis that a is in $\boldsymbol{\mathcal{K}}$ is more favorable and our decision will therefore follow this way. Put in a formal way, we define the function

$$T\ :\ \begin{cases} T(a) = 1 \text{ when } \dim\ker D\tilde{\boldsymbol{P}}_a = t \text{ with } \pi^+(t) \geq \pi^-(t) \\ T(a) = 0 \text{ when } \dim\ker D\tilde{\boldsymbol{P}}_a = t \text{ with } \pi^+(t) \leq \pi^-(t) \end{cases}$$

T is our distinguisher of kernel elements. We next compute its advantage.

Advantage of the Distinguisher The advantage of T for a random instance of the cryptosystem and a random a is by definition

$$|\Pr[T(a) = 1 \mid a \in \mathcal{K}] - \Pr[T(a) = 1 \mid a \notin \mathcal{K}]|$$

The inner difference values to

$$\sum_{t:\pi^+(t) \geq \pi^-(t)} \Pr\left[\dim\ker D\tilde{\boldsymbol{P}}_a = t \mid a \in \mathcal{K}\right] - \Pr\left[\dim\ker D\tilde{\boldsymbol{P}}_a = t \mid a \notin \mathcal{K}\right]$$

The summand of the above is $\pi^+(t) - \pi^-(t)$ and is therefore positive for the prescribed values of t. Hence, the expected advantage of the distinguisher for a random instance of the cryptosystem, denoted Adv, is

$$Adv = \sum_{t:\pi^+(t) \geq \pi^-(t)} \pi^+(t) - \pi^-(t)$$

We summarize in the table below the values of Adv for some parameters.

(n, D, r)	Adv
$(89, 2, 2)$	$2^{-7.49}$
$(89, 3, 2)$	$2^{-12.95}$
$(89, 3, 3)$	$2^{-16.17}$
$(89, 4, 4)$	$2^{-27.97}$

In the above table, the second line corresponds to the preferred parameters in [4].

5 Recovering the Kernel of the Internal Perturbation

In the previous section, we designed a distinguisher T which can be seen as a two-sided error test of membership to $\boldsymbol{\mathcal{K}}$. In this section, we aim at turning the test T into an algorithm for finding elements of $\boldsymbol{\mathcal{K}}$.

5.1 Behaviour of the Test with respect to Linearity

The set $\boldsymbol{\mathcal{K}}$ benefits from a property that its complement does not share: it is closed under addition. Accordingly, when x is a member of $\boldsymbol{\mathcal{K}}$ then any y and $x + y$ must be both members of both non-members of $\boldsymbol{\mathcal{K}}$, while it can happen differently when x is not in $\boldsymbol{\mathcal{K}}$. Analogously, the probability for a random y that both y and $x + y$ are detected inside or outside $\boldsymbol{\mathcal{K}}$ by the test should be higher on average over the elements x of $\boldsymbol{\mathcal{K}}$ than over those not in $\boldsymbol{\mathcal{K}}$. We next show that this intuition is correct and compute the distance between these two probabilities.

Given an element y, we denote by μ_y^+ the probability that $T(x + y) = T(y)$ when x is in $\boldsymbol{\mathcal{K}}$, and by μ_y^- the same probability when x is outside $\boldsymbol{\mathcal{K}}$.

$$\begin{aligned}\mu_y^+ &= \Pr_x[\,T(x + y) = T(y) \mid x \in \boldsymbol{\mathcal{K}}\,]\\ \mu_y^- &= \Pr_x[\,T(x + y) = T(y) \mid x \notin \boldsymbol{\mathcal{K}}\,]\end{aligned}$$

The mean values of μ_y^+ and μ_y^- over the y are denoted μ^+ and μ^-.

$$\mu^+ = \mathrm{Pr}_{x,y}[\,T(x+y) = T(y)\,|\,x \in \boldsymbol{\mathcal{K}}\,]$$
$$\mu^- = \mathrm{Pr}_{x,y}[\,T(x+y) = T(y)\,|\,x \notin \boldsymbol{\mathcal{K}}\,]$$

Probabilities μ^+ and μ^- can be computed for a random instance of the cryptosystem; their distance denoted $\Delta\mu$ is

$$\Delta\mu = \mu^+ - \mu^- = 2.\frac{Adv^2}{2^r} \tag{2}$$

The details of these computations can be found in the full paper.

Given an element y, we define the random variable δ_y which values 1 at x whenever $T(x+y) = T(y)$ and 0 otherwise. The mean value of δ_y over $\boldsymbol{\mathcal{K}}$ is μ_y^+, and is μ_y^- over the complement of $\boldsymbol{\mathcal{K}}$. In the sequel, we will consider a large assembly of random variables δ_{y_i} for some fixed y_i. The idea is that, whenever $\delta_{y_i}(x)$ is 1 for many i, x should belong to $\boldsymbol{\mathcal{K}}$ with high probability.

5.2 Building a Reliable Test of Membership

Definition of the Test For any N non-zero distinct elements $y_1, \ldots, y_N$, we define the random variable

$$S_N(x) = \sum_{i=1}^{N} \delta_{y_i}(x)$$

For any such random variable S_N, a test of membership can be defined as follows. Given an element x, we compute $S_N(x)$; whenever $S_N(x) \geq N\mu^+$, the test answers `yes`, and `no` otherwise.

The intention behind the test is the following. Since $\delta_{y_i}(x)$ is more likely to be 1 when x is in $\boldsymbol{\mathcal{K}}$ than when x is not in $\boldsymbol{\mathcal{K}}$, we expect $S_N(x)$ to be higher when x is in $\boldsymbol{\mathcal{K}}$ than when x is not in $\boldsymbol{\mathcal{K}}$. When N increases, we expect the intersection between the values of S_N over $\boldsymbol{\mathcal{K}}$ and the values of S_N outside $\boldsymbol{\mathcal{K}}$ to become smaller. Finally, for N large enough, we expect the probability that $S_N(x) \geq N\mu^+$ to be large when x is in $\boldsymbol{\mathcal{K}}$ and very small when x is not in $\boldsymbol{\mathcal{K}}$.

Analysis of the Test Let us first consider S_N over $\boldsymbol{\mathcal{K}}$. For any y_i, the mean value of δ_{y_i} over $\boldsymbol{\mathcal{K}}$ is $\mu_{y_i}^+$. This latter value is not known, however we know that it follows a distribution of mean value μ^+. Likewise, the mean value of S_N over $\boldsymbol{\mathcal{K}}$, denoted A_N^+, follows a distribution over the N-tuples $(y_1, \ldots, y_N)$ of mean value $N\mu^+$. Hence, for half the choices of a N-tuple $(y_1, \ldots, y_N)$, we have $A_N^+ \geq N\mu^+$. When this is the case, we have :

$$\mathrm{Pr}_x \left[S_N(x) \geq N\mu^+ \,|\, x \in \boldsymbol{\mathcal{K}} \right] \geq \mathrm{Pr}_x \left[S_N(x) \geq A_N^+ \,|\, x \in \boldsymbol{\mathcal{K}} \right] = \frac{1}{2}$$

Therefore, in at least half the cases, more than the half of the elements of $\boldsymbol{\mathcal{K}}$ will pass our test of membership, whatever is the value of N.

Now we consider S_N over the complement of $\mathcal{K}$. We want to find some N so that the probability for the elements of the complement of $\mathcal{K}$ to pass the test is very small. We can notice as before, that the mean value of S_N over the complement of $\mathcal{K}$, denoted A_N^-, follows a distribution of mean value $N\mu^-$. Hence, for half the choices of a N-tuple $(y_1, \ldots, y_N)$, we have $A_N^- \leq N\mu^-$. When this is the case, we have :

$$\Pr_x \left[S_N(x) \geq N\mu^+ \,|\, x \notin \mathcal{K} \right] \leq \Pr_x \left[S_N(x) - A_N^- \geq N\Delta\mu \,|\, x \notin \mathcal{K} \right] \tag{3}$$

and our task is now to find an upper-bound of the right-hand probability.

We observe that, when the y_i are independently chosen, the random variables δ_{y_i} are independent. Sequences of independent non-identically distributed binary random variables are known as Poisson trials in the litterature. Applying the Chernoff bound [16]:

$$\Pr_x \left[S_N(x) - A_N^- \geq N\Delta\mu \,|\, x \notin \mathcal{K} \right] \leq \exp(-\frac{1}{4}\frac{N^2\Delta\mu^2}{A_N^-})$$

Besides, we have $A_N^- \leq N\mu^-$ and $\mu^- \leq \mu$ where μ is the probability to have $T(x+y) = T(y)$ for random x and y. Therefore, using (3), we finally obtain:

$$\Pr_x \left[S_N(x) \geq N\mu^+ \,|\, x \notin \mathcal{K} \right] \leq \exp(-\frac{N}{4}\frac{\Delta\mu^2}{\mu})$$

We now estimate the value of μ. When x and y are random, $x+y$ and y are independent and therefore $\mu = \alpha^2 + (1-\alpha)^2$ where $\alpha = \Pr[T = 1]$. Probability α can be computed for a random instance of the cryptosystem from

$$\alpha = \sum_{t:\pi^+(t) \geq \pi^-(t)} (2^{-r})\pi^+(t) + (1-2^{-r})\pi^-(t) \simeq \sum_{t:\pi^+(t) \geq \pi^-(t)} \pi^-(t)$$

Using the table 1, we see that $\alpha \simeq 0.6$ and $\mu \simeq 0.5$.

Finally, to make the probability to have a false-positive under ϵ, we can take

$$N = \frac{2}{\Delta\mu^2} \ln\left(\frac{1}{\epsilon}\right) = \frac{2^{2r-1}}{Adv^4} \ln\left(\frac{1}{\epsilon}\right) \tag{4}$$

Complexity for Recovering $\mathcal{K}$ A random element x is in $\mathcal{K}$ with probability $\frac{1}{2^r}$ and is detected in $\mathcal{K}$ by the test with probability $\frac{1}{2}$. Computing all the $\delta_{y_i}(x)$ values is achieved by computing the differentials at $x + y_i$ and at y_i, and then computing their ranks. The complexity for computing a differential or a rank is n^3, the same as for evaluating the public key. Recovering $\mathcal{K}$ requires to discover about n of its elements. Therefore, the complexity for recovering $\mathcal{K}$ is $2^{r+1}Nn$ evaluations of the public key. When taking N as given by Formula 4, recovering $\mathcal{K}$ amounts to

$$\frac{n2^{3r}}{Adv^4} \ln\left(\frac{1}{\epsilon}\right)$$

evaluations of the public key. This is given by the table below for practical parameters and $\epsilon = 0.001$. It should be remarked that Formula 4 gives us an upper-bound on the value of N to be chosen. In practice, taking a smaller N might allow the attack as well.

(n, D, r)	Recovering $\mathcal{K}$
$(89, 2, 1)$	$2^{32.26}$
$(89, 2, 2)$	$2^{45.20}$
$(89, 3, 2)$	$2^{67.03}$
$(89, 3, 3)$	$2^{82.92}$

In the above table, the third line corresponds to the preferred parameters in [4].

6 Invertion of the Public Key

At this point, we assume that $\mathcal{K}$ has been retrieved using the preceding techniques. We next show how the public key of the Internally Perturbed HFE can be inverted using the attack of Faugère-Joux against HFE.

Let $l_1, \ldots, l_r$ to be r independent linear forms orthogonal to $\mathcal{K}$; an element $(x_1, \ldots, x_n)$ lies in $\mathcal{K}$ if and only if $l_k(x_1, \ldots, x_n) = 0$ for all k in $[1, r]$. As already pointed, the public key of an Internally Perturbed HFE is just an HFE public key on any affine subspace parallel to $\mathcal{K}$. Fixing one such subspace, we call $p_1, \ldots, p_n$ the multivariate quadratic forms of the perturbed public key, and $p'_1, \ldots, p'_n$ the multivariate quadratic forms of its equivalent HFE public key on this affine subspace. All linear forms l_k are constant on this subspace; for instance they all value to 0 (the affine subspace considered is $\mathcal{K}$). For any point $(b_1, \ldots, b_n)$, the multivariate quadratic systems $\{p_i = b_i, i \in [1, n]\} \cap \{l_k = 0, k \in [1, r]\}$ and $\{p'_i = b_i, i \in [1, n]\} \cap \{l_k = 0, k \in [1, r]\}$ have the same solutions. Equivalently, the ideal generated by $p_1 - b_1, \ldots, p_n - b_n$ together with $l_1, \ldots, l_r$ is the same as the ideal generated by $p'_1 - b_1, \ldots, p'_n - b_n$ together with $l_1, \ldots, l_r$ in the ring $R = \mathbb{F}_2[x_1, \ldots, x_n]/\{x_1^2 - x_1, \ldots, x_n^2 - x_n\}$. We call I this ideal, and J the ideal generated by $p'_1 - b_1, \ldots, p'_n - b_n$ without the kernel linear forms.

The ideal J is generated by quadratic equations coming from an HFE cryptosystem; computing a Gröbner basis for such ideals was shown much easier than in the general case by Faugère and Joux [6]. In particular, Faugère could break an HFE with parameters $n = 80$ and $D = 6$ in a hundred hours, while HFE arising in practical realizations of the perturbed HFE scheme have suggested parameters $n = 89$ and $D = 3$ only [4]. Now the key point is: computing a Gröbner basis of I cannot be harder than computing a Gröbner basis of J. Indeed I and J only differ by generators of degree 1, and computing a Gröbner basis of these generators is achieved by simple Gaussian elimination. Rather, they will help in the reduction of higher degree polynomials occurring in the computation. This is experimentally checked, as it could be done in about 2h10 when feeding with public and kernel equations and about 2h45 for the corresponding HFE, for any tested instance of the cryptosystem with $(n, D, r) = (60, 3, 2)$, using Magma's implementation of the $F4$ algorithm [18] on a standard machine.

Of course, in practice, $b_1, \dots, b_n$ are made variables and the Gröbner basis computation is made only once. It outputs a set of polynomials $g_1, \dots, g_L$ with the shape, $g_l = f_l(x_1, \dots, x_{i_l}) - h_l(b_1, \dots, b_n)$ where f_l is only in the i_l first x_i. This Gröbner basis allows to solve the system $\{p_1 = b_1, \dots, p_n = b_n\}$ for any values $b_1, \dots, b_n$ by sequentially solving the equations $f_l(x_1, \dots, x_{i_l}) = h_l(b_1, \dots, b_n)$ in increasing order of i_l.

7 Conclusion

The Internally Perturbed HFE cryptosystem is a variation of HFE, designed to fix the potential vulnerability of HFE against algebraic attacks. It is one of the rare candidates liable to enhance HFE as a cryptosystem. A major security element of the cryptosystem is the kernel of the perturbation, since the knowledge of this subspace allows to view the public key as a small set of HFE public keys, which can be inverted for the suggested parameters. However, in this work, we show that some correlation exists between the membership to the kernel of the perturbation and the kernel-dimension of the differential of the public key. This correlation can be accurately measured for any parameters, using sophisticated methods based on combinatorics in binary vector spaces. It yields a distinguisher which can be turned into an algorithm for finding elements of the kernel of the perturbation. For the preferred parameters in [4], recovering the kernel of the perturbation amounts to at most 2^{67} evaluations of the public key, which is well below the usual 2^{80} barrier. Although the designers of the scheme believed that the best attack might be exhaustive search in the space of messages [4], our attack is at least 2^{22} times faster and recovers an equivalent secret key. Accordingly, the elements presented in this work shed a new light on the security of the scheme presented by Ding and Schmidt. It should be emphasized that these elements could not be perceived without the advanced combinatorial methods provided in this paper, which are of independent interest.

References

1. Nicolas Courtois. The Security of Hidden Field Equations (HFE). In David Naccache, editor, *CT-RSA*, volume 2020 of *Lecture Notes in Computer Science*, pages 266–281. Springer, 2001.
2. Jintai Ding. A New Variant of the Matsumoto-Imai Cryptosystem through Perturbation. In Feng Bao, Robert H. Deng, and Jianying Zhou, editors, *Public Key Cryptography*, volume 2947 of *Lecture Notes in Computer Science*, pages 305–318. Springer, 2004.
3. Jintai Ding, Jason E. Gower, Dieter Schmidt, Christopher Wolf, and Z. Yin. Complexity Estimates for the F_4 Attack on the Perturbed Matsumoto-Imai Cryptosystem. In Nigel P. Smart, editor, *IMA Int. Conf.*, volume 3796 of *Lecture Notes in Computer Science*, pages 262–277. Springer, 2005.
4. Jintai Ding and Dieter Schmidt. Cryptanalysis of HFEv and Internal Perturbation of HFE. In Serge Vaudenay, editor, *Public Key Cryptography*, volume 3386 of *Lecture Notes in Computer Science*, pages 288–301. Springer, 2005.

5. Vivien Dubois, Louis Granboulan, and Jacques Stern. An Efficient Provable Distinguisher for HFE. In Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener, editors, *ICALP (2)*, volume 4052 of *Lecture Notes in Computer Science*, pages 156–167. Springer, 2006.
6. Jean-Charles Faugère and Antoine Joux. Algebraic Cryptanalysis of Hidden Field Equation (HFE) Cryptosystems Using Gröbner Bases. In Dan Boneh, editor, *CRYPTO*, volume 2729 of *Lecture Notes in Computer Science*, pages 44–60. Springer, 2003.
7. Harriet J. Fell and Whitfield Diffie. Analysis of a Public Key Approach Based on Polynomial Substitution. In Hugh C. Williams, editor, *CRYPTO*, volume 218 of *Lecture Notes in Computer Science*, pages 340–349. Springer, 1985.
8. Pierre-Alain Fouque, Louis Granboulan, and Jacques Stern. Differential Cryptanalysis for Multivariate Schemes. In Ronald Cramer, editor, *EUROCRYPT*, volume 3494 of *Lecture Notes in Computer Science*, pages 341–353. Springer, 2005.
9. Aviad Kipnis and Adi Shamir. Cryptanalysis of the HFE Public Key Cryptosystem by Relinearization. In Michael J. Wiener, editor, *CRYPTO*, volume 1666 of *Lecture Notes in Computer Science*, pages 19–30. Springer, 1999.
10. Louis Granboulan and Antoine Joux and Jacques Stern. Inverting HFE Is Quasipolynomial. In Cynthia Dwork, editor, *CRYPTO*, volume 4117 of *Lecture Notes in Computer Science*, pages 345–356. Springer, 2006.
11. Tsutomu Matsumoto and Hideki Imai. Public Quadratic Polynominal-Tuples for Efficient Signature-Verification and Message-Encryption. In *EUROCRYPT*, pages 419–453, 1988.
12. M.Garey and D.Johnson. *Computer and Intractability: A guide to the theory of NP-completeness*. Freeman, 1979.
13. NESSIE. European project IST-1999-12324 on New European Schemes for Signature, Integrity and Encryption.
14. Jacques Patarin. Cryptanalysis of the Matsumoto and Imai Public Key Scheme of Eurocrypt'88. In Don Coppersmith, editor, *CRYPTO*, volume 963 of *Lecture Notes in Computer Science*, pages 248–261. Springer, 1995.
15. Jacques Patarin. Hidden Fields Equations (HFE) and Isomorphisms of Polynomials (IP): Two New Families of Asymmetric Algorithms. In *EUROCRYPT*, pages 33–48, 1996.
16. R.Motwani and P.Raghavan. *Randomized Algorithms*, chapter 4, pages 67–74. Cambridge University Press, 1995.
17. Adi Shamir. Efficient Signature Schemes Based on Birational Permutations. In Douglas R. Stinson, editor, *CRYPTO*, volume 773 of *Lecture Notes in Computer Science*, pages 1–12. Springer, 1993.
18. University of Sydney Computational Algebra Group. The MAGMA Computational Algebra System.

A The Kernel-Dimension of the Sum of a Random $\mathbb{F}_2$-Linear Polynomial and a Random Linear Map of Rank r

In characteristic 2, the kernel of the sum of two linear maps is the subspace where they coincide. We denote $\mathcal{L}^D$ the set of $\mathbb{F}_2$-linear polynomials of degree 2^D and $\mathcal{L}_r$ the set of linear maps of rank r. We aim at determining the distribution of probability of the dimension of the subspace where L and l coincide, denoted $\{L = l\}$, when L is a random element of $\mathcal{L}^D$ and l is a random element of $\mathcal{L}_r$.

We recall that the number $S(n, s)$ of linearly independent sequences of length s in a space of dimension n is

$$S(n, s) = \prod_{i=0}^{s-1} (2^n - 2^i)$$

Each such sequence generates a subspace of dimension s which is also generated by $S(s, s)$ other linearly independent sequences of length s. Therefore the number $E(n, s)$ of subspaces of dimension s in a space of dimension n is $S(n, s)/S(s, s)$.

An $\mathbb{F}_2$-linear polynomial of degree 2^D has at most 2^D roots as a polynomial. Its roots are the elements of its kernel as a linear map. Therefore the dimension of this kernel cannot exceed D and the probability that it has kernel dimension d is given by the following lemma:

Lemma 1. *The probabilities $(p_D(0), \ldots, p_D(D))$ that a random element of $\mathcal{L}^D$ has kernel dimension respectively $0, \ldots, D$ satisfy the following invertible triangular system:*

$$d \in [0, D], \quad E(n, d)2^{-nd} = \sum_{m=d}^{D} E(m, d) p_D(m)$$

Proof. The number of $\mathbb{F}_2$-linear polynomials of degree 2^D is $(2^n - 1)2^{nD}$ and those which vanish at a are 2^n times less numerous. Given a subspace of dimension d with d in $[0, D]$, the vanishing of an $\mathbb{F}_2$-linear polynomial of degree 2^D results in d linear constraints over its $D + 1$ coefficients. It implies that for each subspace of dimension d, there are exactly $(2^n - 1)2^{n(D-d)}$ $\mathbb{F}_2$-linear polynomials which vanish on it. In the product $E(n, d)(2^n - 1)2^{n(D-d)}$, the $\mathbb{F}_2$-linear polynomials whose kernel has dimension m with $m \geq d$ are counted $E(m, d)$ times. Therefore, the proportions $p_D(d)$ of $\mathbb{F}_2$-linear polynomials of degree 2^D which have kernel dimension d satisfy the above invertible triangular system. □

We now suppose given an $\mathbb{F}_2$-linear polynomial L of degree 2^D and kernel-dimension d. The subspace on which L and a randomly chosen linear map of rank r coincide at least contains the intersection of the two kernels. We therefore should fix this dimension of intersection as a new parameter.

Lemma 2. *Given a subspace of dimension d, the probability $p_{d,r}(i)$ that a random subspace of dimension $n - r$ intersects this subspace with dimension i is*

$$\frac{S(d, i)S(n, d + n - r - i)S(n - r, i)}{S(n, d)S(i, i)S(n, n - r)}$$

Proof. Let call F the prescribed subspace of dimension d. The number of possible intersection subspaces is $E(d, i)$. For each of them I, the number of linearly independent sequences of length $n-r$ whose generating subspace has intersection with F exactly I is the number of linearly independent sequences outside F:

$$(2^n - 2^d)\dots(2^n - 2^{d+n-r-i-1}) = S(n, d+n-r-i)/S(n,d)$$

This generating subspace G is also generated by as many linearly independent sequences of length $n-r$ as the number of linearly independent sequences of length $n-r-i$ of G outside I; this is likewise $S(n-r, n-r)/S(n-r, i)$.

The number of subspaces of dimension $n-r$ which intersect F with dimension i is therefore

$$E(d,i)\frac{S(n,d+n-r-i)}{S(n,d)}\frac{S(n-r,i)}{S(n-r,n-r)}$$

and the expected proportion is obtained by dividing by $E(n, n-r)$. $\square$

We now suppose given both an $\mathbb{F}_2$-linear polynomial L of degree 2^D and kernel F of dimension d and a subspace G of dimension $n-r$ which has intersection of dimension i with F. A map of kernel G coincides with L on a subspace H such that $H \cap F = H \cap G = F \cap G$. We now enumerate the number of subspaces of dimension t satisfying this condition.

Lemma 3. *Given a subspace F of dimension d and a subspace G of dimension $n-r$ whose intersection has dimension i, the number of subspaces of dimension t such that $H \cap F = H \cap G = F \cap G$ is*

$$\sum_{j=i}^{t} \frac{S(d-i,j-i)S(n-r-i,j-i)}{S(j-i,j-i)}\frac{S(n,d+n-r-i+t-j)}{S(n,d+n-r-i)}\frac{S(t,j)}{S(t,t)}$$

Proof. This enumeration comes in two steps: first we count the number of subspaces J of $F+G$ which have dimension j and satisfy the condition, second we count for each such J the number of subspaces H of dimension t whose intersection with $F+G$ is J.

The subspaces of $F+G$ of dimension j containing $F \cap G$ are in bijection with the subspaces of dimension $j-i$ in the quotient space $(F+G)/(F\cap G)$. Let $\bar{x}$ denote the class modulo $F \cap G$ of the element x. The number of subspaces J such that $F \cap J = G \cap J = F \cap G$ is the number of subspaces $\bar{J}$ such that $\overline{F} \cap \bar{J} = \bar{G} \cap \bar{J} = \{\bar{0}\}$ in the quotient space. Now notice that the set of linearly independent sequences of length $j-i$ in $\overline{F} + \bar{G}$ generating a subspace of zero intersection with both $\overline{F}$ and $\bar{G}$ is in bijection with the Cartesian product of lin. indep. sequences of length $j-i$ in $\overline{F}$ and lin. indep. sequences of length $j-i$ in $\bar{G}$. Besides each such sequence generates a subspace which is also generated by $S(j-i, j-i)$ others. The number of subspaces J of $F+G$ of dimension j such that $J \cap F = J \cap G = F \cap G$ is therefore $S(d-i,j-i)S(n-r-i,j-i)/S(j-i,j-i)$.

The number of subspaces of dimension t whose intersection with $F+G$ has dimension j is enumerated as given by Lemma 2. $\square$

It now only remains to determine the proportion of linear maps of kernel G which coincide with L on a subspace of dimension t.

Lemma 4. *Let L a linear map of kernel F of dimension d, G a subspace of dimension $n-r$ which has intersection of dimension i with F and $E_{d,r,i}(t)$ the number of subspaces H of dimension t such that $H \cap F = H \cap G = F \cap G$. The proportions $p_{d,r,i}(t)$ of linear maps of kernel G which coincide with L on a subspace of dimension t for t in $[i, r+i]$ satisfy the following invertible triangular system:*

$$t \in [i, r+i], \quad \frac{E_{d,r,i}(t)}{S(n, t-i)} = \sum_{m=t}^{r+i} E(m,t) \frac{S(t,i)}{S(m,i)} p_{d,r,i}(m)$$

Proof. For each subspace H of dimension t such that $H \cap F = H \cap G = F \cap G$, we construct a linear map of kernel G which equal L on H by choosing for its image on the remaining dimension $r-t+i$ a linearly independent sequence outside the image of H by L which has dimension $t-i$. The number of such maps is thus $S(n,r)/S(n,t-i)$, and their proportion over all maps of kernel G is $1/S(n,t-i)$. Now, making the product of the number of subspaces H of dimension t and satisfying $H \cap F = H \cap G = F \cap G$ by the number of linear maps l of kernel G which coincide with L on H, we see that the linear maps of kernel G which coincide with L on a subspace of dimension $m \geq t$ are counted as many as the number of subspaces of dimension t containing $F \cap G$ in this subspace. This number is $E(m,t)S(t,i)/S(m,i)$ as it can be easily checked. $\square$

Putting all this together, we obtain that the probability that a random $\mathbb{F}_2$-linear polynomial L of degree 2^D and kernel F of dimension d coincides on a subspace of dimension t with a linear map l of rank r whose kernel has intersection of dimension i with F is

$$\pi_{d,r,i}(t) = p_D(d) p_{d,r}(i) p_{d,r,i}(t)$$

Of course the probability in term of the sole parameter t comes by summing over all possible values for d and i.