A Proof of Security in $O(2^n)$ for the Xor of Two Random Permutations
– Proof with the "H_σ technique"–

Jacques Patarin
Université de Versailles
45 avenue des Etats-Unis
78035 Versailles Cedex - France

Abstract

Xoring two permutations is a very simple way to construct pseudorandom functions from pseudorandom permutations. The aim of this paper is to get precise security results for this construction. Since such construction has many applications in cryptography (see [2, 3, 4, 6] for example), this problem is interesting both from a theoretical and from a practical point of view. In [6], it was proved that Xoring two random permutations gives a secure pseudorandom function if $m \ll 2^{\frac{2n}{3}}$. By "secure" we mean here that the scheme will resist all adaptive chosen plaintext attacks limited to m queries (even with unlimited computing power). More generally in [6] it is also proved that with k Xor, instead of 2, we have security when $m \ll 2^{\frac{kn}{k+1}}$. In this paper we will prove that for $k = 2$, we have in fact already security when $m \ll O(2^n)$. Therefore we will obtain a proof of a similar result claimed in [2] (security when $m \ll O(2^n/n^{2/3})$). Moreover our proof is very different from the proof strategy suggested in [2] (we do not use Azuma inequality and Chernoff bounds for example, but we will use the "H_σ technique" as we will explain), and we will get precise and explicit O functions. Another interesting point of our proof is that we will show that this (cryptographic) problem of security is directly related to a very simple to describe and purely combinatorial problem.

Key words: Pseudorandom functions, pseudorandom permutations, security beyond the birthday bound, Luby-Rackoff backwards,H_σ technique, introduction to Mirror Theory.

This paper is the extended version of the paper [15] with the same title published at ICITS 2008 pp. 232-248.

1 Introduction

The problem of converting pseudorandom permutations (PRP) into pseudorandom functions (PRF) named "Luby-Rackoff backwards" was first considered in [3]. This problem is obvious if we are interested in an asymptotical polynomial versus non polynomial security model (since a PRP is then a PRF), but not if we are interested in achieving more optimal and concrete security bounds. More precisely, the loss of security when regarding a PRP as a PRF comes from the "birthday attack" which can distinguish a random permutation from a random function of n bits to n bits, in $2^{\frac{n}{2}}$ operations and $2^{\frac{n}{2}}$ queries. Therefore different ways to build PRF from PRP with a security above $2^{\frac{n}{2}}$ and by performing very few computations have been suggested (see [2, 3, 4, 6]). One of the simplest way is simply to Xor k independent pseudorandom permutations, for example with $k = 2$. In [6] (Theorem 2 p.474), it has been proved, with a simple proof, that the Xor of k independent PRP gives a PRF with security at least in $O(2^{\frac{k}{k+1}n})$. (For $k = 2$ this gives $O(2^{\frac{2}{3}n})$). In [2], a much more complex strategy (based on Azuma inequality and Chernoff bounds) is presented. It is claimed that with this strategy we may prove that the Xor of two PRP gives a PRF with security at least in $O(2^n/n^{\frac{2}{3}})$ and at most in $O(2^n)$, which is much better than the birthday bound in $O(2^{\frac{n}{2}})$. However the authors of [2] present a very general framework of proof and they do not give every details for this result. For example, page 9 they wrote "we give only a very brief summary of how this works", and page 10 they introduce O functions that are not easy to express explicitly. In this paper we will use a completely different proof strategy, based on the "H_σ technique" (this is part of the general "coefficient H technique", see Section 3 below), simple counting arguments and induction. We will need a few pages, but we will get like this a self contained proof of security in $O(2^n)$ for the Xor of two permutations with a very precise O function. In fact, this paper can be seen as a good introduction to this "H_σ technique". (This technique can also be used for the proof of many other secret key schemes). Since building PRF from PRP has many applications (see [2, 3, 4]), we think that these results are really interesting both from theoretical and from practical point of view.

It may be also interesting to notice that there are many similarities between this problem and the security of Feistel schemes built with random round functions (also called Luby-Rackoff constructions). In [8], it was proved that for L-R constructions with k rounds functions we have security that tends to $O(2^n)$ when the number k of rounds tends to infinity. Then in [13], it was proved that security in $O(2^n)$ was obtained not only for $k \to +\infty$, but already for $k = 7$ (Later similar proofs for $k = 6$ and $k = 5$ were obtained). Similarly, we have seen that in [6] it was proved that for the Xor of k PRP we have security that tends $O(2^n)$ when $k \to +\infty$. In this paper, we show that security in $O(2^n)$ is not only for $k \to +\infty$, but already for $k = 2$.

Related Problems. In [9] the best know attacks on the Xor of k random permutations are studied in various scenarios. For $k = 2$ the bound obtained are near our security bounds. In [7] attacks on the Xor of two **public** permutations are studied (i.e. indifferentiability instead of indistinguishibility).

Part I

From the Xor of Two Permutations to the λ_i values

2 Notation and Aim of this paper

In all this paper we will denote $I_n = \{0,1\}^n$. F_n will be the set of all applications from I_n to I_n, and B_n will be the set of all permutations from I_n to I_n. Therefore $|I_n| = 2^n$, $|F_n| = 2^{n \cdot 2^n}$ and $|B_n| = (2^n)!$. $x \in_R A$ means that x is randomly chosen in A with a uniform distribution.

The aim of this paper is to prove the theorem below, with an explicit O function (to be determined).

Theorem 1 *For all CPA-2 (Adaptive chosen plaintext attack) ϕ on a function G of F_n with m chosen plaintext, we have:* $\mathrm{Adv}_\phi^{\mathrm{PRF}} \le O(\frac{m}{2^n})$ *where* $\mathrm{Adv}_\phi^{\mathrm{PRF}}$ *denotes the advantage to distinguish $f \oplus g$, with $f, g \in_R B_n$ from $h \in_R F_n$.*

This theorem says that there is no way (with an adaptive chosen plaintext attack) to distinguish with a good probability $f \oplus g$ when $f, g \in_R B_n$ from $h \in_R F_n$ when $m \ll 2^n$ (and this even if we have access to infinite computing power, as long as we have access to only m queries). Therefore, it implies that the number λ of computations to distinguish $f \oplus g$ with $f, g \in_R B_n$ from $h \in_R F_n$ satisfies: $\lambda \geq O(2^n)$. We say also that there is no generic CPA-2 attack with less than $O(2^n)$ computations for this problem, or that the security obtained is greater than or equal to $O(2^n)$. Since we know (for example from [2] or Attack 1 of Appendix F) that there is an attack in $O(2^n)$, Theorem 1 also says that $O(2^n)$ is the exact security bound for this problem.

3 The general Proof Strategy ("H_σ technique")

We will use this general Theorem:

Theorem 2 *("Coefficient H technique") Let α and β be real numbers, $\alpha > 0$ and $\beta > 0$. Let $\mathcal{E}$ be a subset of I_n^m such that $|\mathcal{E}| \geq (1-\beta) \cdot 2^{nm}$. If:*

1. *For all sequences a_i, $1 \leq i \leq m$, of pairwise distinct elements of I_n and for all sequences b_i, $1 \leq i \leq m$, of $\mathcal{E}$ we have:*

$$H \geq \frac{|B_n|^2}{2^{nm}}(1-\alpha)$$

where H denotes the number of $(f,g) \in B_n^2$ such that $\forall i,\ 1 \leq i \leq m,\ (f \oplus g)(a_i) = b_i$.

Then

2. *For every CPA-2 with m chosen plaintexts we have: $p \leq \alpha + \beta$ where $p = \mathrm{Adv}_\phi^{\mathrm{PRF}}$ denotes the advantage to distinguish $f \oplus g$ when $(f,g) \in_R B_n^2$ from a function $h \in_R F_n$.*

Remark. H is a simplified notation for $H(a,b)$, or for $H(b)$ since we can easily prove that $H(a,b)$ does not depend of the $a = (a_i,\ 1 \leq i \leq m)$ values (but in general depends of the $b = (b_i,\ 1 \leq i \leq m)$ values).
Proof: Let a'_i, $1 \leq i \leq m$ be a sequence of pairwise distinct elements of I_n and let φ be a bijection such that $\forall i,\ 1 \leq i \leq m,\ \varphi(a'_i) = a_i$. Then: $f \circ \varphi(a'_i) \oplus g \circ \varphi(a'_i) = b_i \Leftrightarrow f(a_i) \oplus g(a_i) = b_i$. Thus we see that $H(a'_i, b'_i) \geq H(a_i, b_i)$ and similarly $H(a_i, b_i) \leq H(a'_i, b_i)$.

Proof of Theorem 2

It is not very difficult to prove Theorem 2 with classical counting arguments. This proof technique is sometimes called the "Coefficient H technique". A complete proof of Theorem 2 can also be found in [12] page 27 and a similar Theorem was used in [13] p.517. In order to have all the proofs in this paper, Theorem 2 is also proved in Appendix F.

How to get Theorem 1 from Theorem 2("H_σ technique")

In order to get Theorem 1 from Theorem 2, a sufficient condition is to prove that for " most" (most since we need β small) sequences of values b_i, $1 \leq i \leq m$, $b_i \in I_n$, we have: the number H of $(f,g) \in B_n^2$ such that $\forall i,\ 1 \leq i \leq m,\ f(a_i) \oplus g(a_i) = b_i$ satisfies: $H \geq \dfrac{|B_n|^2}{2^{nm}}(1-\alpha)$ for a small value α (more precisely with $\alpha \ll O(\frac{m}{2^n})$). For this, we will evaluate $E(H)$ the mean value of H when the b_i values are randomly chosen in I_n^m, and $\sigma(H)$ the standard deviation of H when the b_i values are randomly chosen in I_n^m. (Therefore we can call our general proof strategy the "Hσ technique", since we use the coefficient H technique plus the evaluation of $\sigma(H)$). We will prove that $E(H) = \frac{|B_n|^2}{2^{nm}}$ and that $\sigma(H) = \frac{|B_n|^2}{2^{nm}} O(\frac{m}{2^n})^{\frac{3}{2}}$, with an explicit O function, i.e. that $\sigma(H) \ll E(H)$ when $m \ll 2^n$. From Bienayme-Tchebichev Theorem, we have

$$\forall \epsilon > 0,\ Pr(|H - E(H)| \leq \epsilon) \geq 1 - \frac{V(H)}{\epsilon^2}$$

So with $\epsilon = \alpha E(H)$, we get:

$$\forall \alpha > 0,\ Pr\big(|H - E(H)| \leq \alpha E(H)\big) \geq 1 - \frac{\sigma^2(H)}{\alpha^2 E^2(H)}$$

So

$$\forall \alpha > 0,\ Pr\big[H \geq E(H)(1-\alpha)\big] \geq 1 - \frac{\sigma^2(H)}{\alpha^2 E^2(H)}$$

Therefore with $\mathcal{E} = \{b_i,\ H(b_i) \geq E(H)(1-\alpha)\}$ from Theorem 2 we will have for all $\alpha > 0$:

$$\text{Adv}_\phi^{\text{PRF}} \leq \alpha + \frac{\sigma^2(H)}{\alpha^2 E^2(H)}$$

With $\alpha = \big(\frac{\sigma(H)}{E(H)}\big)^{2/3}$, this gives

$$\text{Adv}_\phi^{\text{PRF}} \leq 2\big(\frac{\sigma(H)}{E(H)}\big)^{2/3} = 2\big(\frac{V(H)}{E^2(H)}\big)^{1/3} \quad (3.1)$$

So if $\frac{\sigma(H)}{E(H)} = O(\frac{m}{2^n})^{3/2}$, and $E(H) = \frac{|B_n|^2}{2^{nm}}$, Theorem 1 comes from Theorem 2.

Introducing h instead of H

H is (by definition) the number of $(f,g) \in B_n^2$ such that $\forall i,\ 1 \leq i \leq m,\ f(a_i) \oplus g(a_i) = b_i$. $\forall i,\ 1 \leq i \leq m$, let $x_i = f(a_i)$. We will denote $h(b)$, or simply by h, for simplicity (but h depends on b), be the number of sequences x_i, $1 \leq i \leq m$, $x_i \in I_n$, such that:

1. The x_i are pairwise distinct, $1 \leq i \leq m$.
2. The $x_i \oplus b_i$ are pairwise distinct, $1 \leq i \leq m$.
   We see that $H = h \cdot \frac{|B_n|^2}{\big(2^n(2^n-1)\dots(2^n-m+1)\big)^2}$ $(\sharp)$. (Since when x_i is fixed, f and g are fixed on exactly m pairwise distinct points by $\forall i,\ 1 \leq i \leq m$, $f(a_i) = x_i$ and $g(a_i) = b_i \oplus x_i$). ($\sharp$ gives another proof that $H(a,b)$ does not depend on the a_i values).

Thus we have

$$\text{Adv}_\phi^{\text{PRF}} \leq 2\big(\frac{\sigma(H)}{E(H)}\big)^{2/3} = 2\big(\frac{\sigma(h)}{E(h)}\big)^{2/3} \quad (3.1)$$

Therefore, instead of evaluating $E(H)$ and $\sigma(H)$, we can evaluate $E(h)$ and $\sigma(h)$, and our aim is to prove that

$$E(h) = \frac{(2^n(2^n-1)\dots(2^n-m+1))^2}{2^{nm}} \quad \text{(this means that} E(h) = \frac{|B_n|^2}{2^{nm}} \quad \text{from } (\sharp))$$

and that

$$\sigma(h) \ll E(h) \quad \text{when } m \ll 2^n$$

As we will see, the most difficult part will be the evaluation of $\sigma(N)$. (We will see in Section 5 that this evaluation of $\sigma(h)$ leads us to a purely combinatorial problem: the evaluation of values that we will call λ_α).

Remark: We will not do it, nor need it, in this paper, but it is possible to improve slightly the bounds by using a more precise evaluation than the Bienayme-Tchebichev Theorem: instead of

$$Pr(|h - E(h)| \geq t\sigma(h)) \leq \frac{1}{t^2},$$

it is possible to prove that for our variables h, and for $t >> 1$, we have something like this:

$$Pr(|h - E(h)| \geq t\sigma(h)) \leq \frac{1}{\mathrm{e}^t}$$

(For this we would have to analyze more precisely the law of distribution of h: it follows almost a Gaussian and this gives a better evaluation than just the general $\frac{1}{t^2}$).

4 Computation of $E(h)$

Let $b = (b_1, \dots, b_n)$, and $x = (x_1, \dots, x_n)$. For $x \in I_n^m$, let

$$\delta_x = 1 \Leftrightarrow \begin{cases} \text{The } x_i \text{ are pairwise distinct,} & 1 \leq i \leq m \\ \text{The } x_i \oplus b_i \text{ are pairwise distinct,} & 1 \leq i \leq m \end{cases}$$

and $\delta_x = 0 \Leftrightarrow \delta_x \neq 1$. Let J_n^m be the set of all sequences x_i such that all the x_i are pairwise distinct, $1 \leq i \leq m$. Then $|J_n^m| = 2^n(2^n-1)\dots(2^n-m+1)$ and $h = \sum_{x \in J_n^m} \delta_x$. So we have $E(h) = \sum_{x \in J_n^m} E(\delta_x)$. For $x \in J_n^m$,

$$E(\delta_x) = Pr_{b \in_R I_n^m}(\text{All the } x_i \oplus b_i \text{ are pairwise distinct}) = \frac{2^n(2^n-1)\dots(2^n-m+1)}{2^{nm}}$$

Therefore

$$E(h) = |J_n^m| \cdot \frac{2^n(2^n-1)\dots(2^n-m+1)}{2^{nm}} = \frac{(2^n(2^n-1)\dots(2^n-m+1))^2}{2^{nm}}$$

as expected.

5 First results on $V(h)$

We denote by $V(h)$ the variance of h when $b \in_R I_n^m$. We have seen that our aim (cf(3.1)) is to prove that $V(h) \ll E^2(h)$ when $m \ll 2^n$ (with $E^2(h) = \frac{(2^n(2^n-1)\dots(2^n-m+1))^4}{2^{2nm}}$). With the same notations as in Section 4 above, $h = \sum_{x \in J_n^m} \delta_x$. Since the variance of a sum is the sum of the variances plus the sum of all covariances we have:

$$V(h) = \sum_{x,x' \in J_n^m} \left[E(\delta_x\, \delta_{x'}) - E(\delta_x)\, E(\delta_{x'})\right] \quad (5.1)$$

We will now study the 2 terms in (5.1), i.e. the terms in $E(\delta_x\, \delta_{x'})$ and the terms in $E(\delta_x)\, E(\delta_{x'})$.

Terms in $E(\delta_x)\, E(\delta_{x'})$

$$E(\delta_x)\, E(\delta_{x'}) = \frac{(2^n(2^n-1)\dots(2^n-m+1))^2}{2^{2nm}}$$

$$\text{So} \quad \sum_{x,x' \in J_n^m} E(\delta_x) E(\delta_{x'}) = \frac{(2^n(2^n-1)\dots(2^n-m+1))^4}{2^{2nm}} = E^2(N) \quad (5.3)$$

Terms in $E(\delta_x\, \delta_{x'})$

Therefore the last term A_m that we have to evaluate in (5.1) is

$$A_m =_{def} \sum_{x,x' \in J_n^m} E(\delta_x\, \delta_{x'}) =$$

$$\sum_{x,x' \in J_n^m} Pr_{b \in_R I_n^m} \left(\begin{cases} \text{The } x_i \text{ are pairwise distinct,} & 1 \le i \le m \\ \text{The } x'_i \text{ are pairwise distinct,} & 1 \le i \le m \\ \text{The } x_i \oplus b_i \text{ are pairwise distinct,} & 1 \le i \le m \\ \text{The } x'_i \oplus b_i \text{ are pairwise distinct,} & 1 \le i \le m \end{cases} \right)$$

Let $\lambda_m =_{def}$ the number of sequences (x_i, x'_i, b_i), $1 \le i \le m$ such that

1. The x_i are pairwise distinct, $1 \le i \le m$.
2. The x'_i are pairwise distinct, $1 \le i \le m$.
3. The $x_i \oplus b_i$ are pairwise distinct, $1 \le i \le m$.
4. The $x'_i \oplus b_i$ are pairwise distinct, $1 \le i \le m$.

We have $A_m = \frac{\lambda_m}{2^{nm}}$ (5.4). We also have

$$\lambda_m = \sum_{b \in I_n^m} [\text{ Number of sequences } x_i,\ 1 \le i \le m, \text{ such that the } x_i \text{ are pairwise distinct,}$$

$$\text{and the } x_i \oplus b_i \text{ are pairwise distinct }]^2$$

Let $U_m = \frac{(2^n(2^n-1)\dots(2^n-m+1))^4}{2^{nm}} = E^2(h) \cdot 2^{nm}$.

From (5.1), (5.2), (5.3), (5.4), we have obtained:

$$V(h) = \frac{\lambda_m}{2^{nm}} - E^2(h) = \frac{\lambda_m - U_m}{2^{nm}} \quad (5.5)$$

Moreover, from (3.1), we have

$$Adv_\phi^{PRF} \le 2\left(\frac{\lambda_m}{U_m} - 1\right)^{1/3} \quad (5.6)$$

Therefore, our aim is to prove that $\lambda_m \lesssim U_m$

$$\text{i.e. } \lambda_m \lesssim 2^{nm} \cdot E^2(h) = \frac{(2^n(2^n-1)\dots(2^n-m+1))^4}{2^{nm}} \quad (5.7)$$

where $a \lesssim b$ means $a \leq b$ or $a \simeq b$.
Remark. Since $V(h) \geq 0$, we necessarily have from (5.5):

$$\lambda_m \geq U_m, \quad \text{i.e. } \lambda_m \geq E^2(h) \cdot 2^{nm} \quad (5.8)$$

Unfortunately our aim is to prove the other direction: $\lambda_m \lesssim E^2(h) \cdot 2^{nm}$ (it is more difficult). However since we have (5.7) we can notice that proving $\lambda_m \lesssim U_m$ is in fact equivalent to prove $\lambda_m \simeq U_m$. It is interesting to notice that the cryptographic property that we want to prove is "just" equivalent to $\lambda_m \simeq E^2(h) \cdot 2^{nm}$ where the λ_m values do not depend on a or b but only on m. It is also interesting to notice that in "standard " coefficients H theorems we usually want to prove that $H \geq$ something, while here we want to prove that $\lambda_m \leq$ something (by using $\sigma(H)$ instead of H).

Change of variables

Let $f_i = x_i$ and $g_i = x'_i$, $h_i = x_i \oplus b_i$. We see that λ_m is also the number of sequences (f_i, g_i, h_i), $1 \leq i \leq m$, $f_i \in I_n$, $g_i \in I_n$, $h_i \in I_n$, such that

1. The f_i are pairwise distinct, $1 \leq i \leq m$.
2. The g_i are pairwise distinct, $1 \leq i \leq m$.
3. The h_i are pairwise distinct, $1 \leq i \leq m$.
4. The $f_i \oplus g_i \oplus h_i$ are pairwise distinct, $1 \leq i \leq m$.

(With this representation we can express λ_m without introducing the b_i values).
We will call these conditions 1.2.3.4. the "conditions λ_α". (Examples of λ_m values are given in Appendix A). In order to get (5.7), we see that a sufficient condition is finally to prove that

$$\lambda_m \leq \frac{(2^n(2^n-1)\dots(2^n-m+1))^4}{2^{nm}}\big(1+O(\frac{m}{2^n})\big) \quad (5.9)$$

(or $=$ instead of $\leq$ here) with an explicit O function. So we have transformed our security proof against all CPA-2 for $f \oplus g$, $f, g \in_R B_n$, to this purely combinatorial problem (5.9) on the λ_m values. (We can notice that in $E(h)$ and $\sigma(h)$ we evaluate the values when the b_i values are randomly chosen, while here, on the λ_m values, we do not have such b_i values anymore). The proof of this combinatorial property is given below and in the Appendices. (Unfortunately the proof of this combinatorial property (5.9) is not obvious: we will need a few pages. However, fortunately, the mathematics that we will use are simple).
Notation. We will sometime use the notation: $z_i = f_i \oplus g_i \oplus h_i$. Then we can notice that in all our systems the variables f_i, g_i, h_i and z_i are symmetrical, i.e. they have the same properties. Moreover, we can notice that if we remove the equation $z_i = f_i \oplus g_i \oplus h_i$ but keep the fact that $z_i \neq z_j$ if $i \neq j$, then we get exactly $(2^n(2^n-1)\dots(2^n-m+1))^4$ solutions.

Part II
First analysis of the λ_i values, security in $m \ll 2^{\frac{5n}{6}}$

6 First results in λ_α

The values λ_α have been introduced in Section 5. Our aim is to prove (5.9), (or something similar, for example with $O(\frac{m^{k+1}}{2^{nk}})$ for any integer k) with explicit O functions. For this, we will proceed like this: in this Section 6 we will give a first evaluation of the values λ_α. Then, in Section 7, we will prove an induction formula (7.2) on λ_α. Finally, we will use this induction formula (7.2) to get our property on λ_α.

We have defined above: $U_\alpha = \dfrac{[2^n(2^n-1)\dots(2^n-\alpha+1)]^4}{2^{n\alpha}}$. We have $U_{\alpha+1} = \frac{(2^n-\alpha)^4}{2^n}U_\alpha$.

$$U_{\alpha+1} = 2^{3n}\big(1 - \frac{4\alpha}{2^n} + \frac{6\alpha^2}{2^{2n}} - \frac{4\alpha^3}{2^{3n}} + \frac{\alpha^4}{2^{4n}}\big)U_\alpha \quad (6.1)$$

Similarly, we want to obtain an induction formula on λ_α, i.e. we want to evaluate $\frac{\lambda_{\alpha+1}}{\lambda_\alpha}$. More precisely our aim is to prove something like this:

$$\frac{\lambda_{\alpha+1}}{\lambda_\alpha} = \frac{U_{\alpha+1}}{U_\alpha}\big(1 + O(\frac{1}{2^n}) + O(\frac{\alpha}{2^{2n}})\big) \quad (6.2)$$

Notice that here we have $O(\frac{\alpha}{2^{2n}})$ and not $O(\frac{\alpha}{2^n})$. Therefore we want something like this:

$$\underline{\frac{\lambda_{\alpha+1}}{2^{3n}\cdot\lambda_\alpha} = \big(1 - \frac{4\alpha}{2^n} + \frac{6\alpha^2}{2^{2n}} - \frac{4\alpha^3}{2^{3n}} + \frac{\alpha^4}{2^{4n}}\big)\big(1 + O(\frac{1}{2^n}) + O(\frac{\alpha}{2^{2n}})\big)} \quad (6.3)$$

(with some specific O functions)

Then, from (6.2) used for all $1 \le i \le \alpha$ and since $\lambda_1 = U_1 = 2^{3n}$, we will get

$$\lambda_\alpha = \big(\frac{\lambda_\alpha}{\lambda_{\alpha-1}}\big)\big(\frac{\lambda_{\alpha-1}}{\lambda_{\alpha-2}}\big)\dots\big(\frac{\lambda_2}{\lambda_1}\big)\lambda_1 = U_\alpha\big(1 + O(\frac{1}{2^n}) + O(\frac{\alpha}{2^{2n}})\big)^\alpha$$

From this we will get:

$$\lambda_\alpha = U_\alpha\big(1 + O(\frac{\alpha}{2^n})\big)$$

and therefore we will get property (5.9):

$$\lambda_\alpha \le U_\alpha(1 + O(\frac{\alpha}{2^n}))$$

as wanted. Notice that to get here $O(\frac{\alpha}{2^n})$ we have used $O(\frac{\alpha}{2^{2n}})$ in (6.2).

By definition $\lambda_{\alpha+1}$ is the number of sequences (f_i, g_i, h_i), $1 \le i \le \alpha+1$ such that we have:

1. The conditions λ_α
2. $f_{\alpha+1} \notin \{f_1, \dots, f_\alpha\}$
3. $g_{\alpha+1} \notin \{g_1, \dots, g_\alpha\}$

4. $h_{\alpha+1} \notin \{h_1, \ldots, h_\alpha\}$

5. $f_{\alpha+1} \oplus g_{\alpha+1} \oplus h_{\alpha+1} \notin \{f_1 \oplus g_1 \oplus h_1, \ldots, f_\alpha \oplus g_\alpha \oplus h_\alpha\}$

We will denote by $\beta_1, \ldots, \beta_{4\alpha}$ the 4α equalities that should not be satisfied here: $\beta_1 : f_{\alpha+1} = f_1$, $\beta_2 : f_{\alpha+1} = f_2$, …, $\beta_{4\alpha} : f_{\alpha+1} \oplus g_{\alpha+1} \oplus h_{\alpha+1} = f_\alpha \oplus g_\alpha \oplus h_\alpha$.

First evaluation

When f_i, g_i, h_i values are fixed, $1 \leq i \leq \alpha$, such that they satisfy conditions λ_α, for $f_{\alpha+1}$ that satisfy 2), we have $2^n - \alpha$ solutions and for $g_{\alpha+1}$ that satisfy 3) we have $2^n - \alpha$ solutions. Now when f_i, g_i, h_i, $1 \leq i \leq \alpha$, and $f_{\alpha+1}$, $g_{\alpha+1}$ are fixed such that they satisfy 1), 2), 3), for $h_{\alpha+1}$ that satisfy 4) and 5) we have between $2^n - \alpha$ and $2^n - 2\alpha$ possibilities. Therefore (first evaluation for $\frac{\lambda_{\alpha+1}}{\lambda_\alpha}$) we have:

$$\lambda_\alpha (2^n - \alpha)^2 (2^n - 2\alpha) \leq \lambda_{\alpha+1} \leq \lambda_\alpha (2^n - \alpha)^2 (2^n - \alpha)$$

Therefore :

$$1 - \frac{4\alpha}{2^n} + \frac{5\alpha^2}{2^{2n}} - \frac{2\alpha^3}{2^{3n}} \leq \frac{\lambda_{\alpha+1}}{2^{3n} \cdot \lambda_\alpha} \leq 1 - \frac{3\alpha}{2^n} + \frac{3\alpha^2}{2^{2n}} - \frac{\alpha^3}{2^{3n}} \leq 1 \quad (6.4)$$

This is an approximation in $O(\frac{\alpha}{2^n})$ and from it and (6.1) we get

$$\lambda_\alpha = U_\alpha \big(1 + O(\frac{\alpha}{2^n})\big)^\alpha$$

$$\text{i.e.} \quad \lambda_\alpha = U_\alpha \big(1 + O(\frac{\alpha^2}{2^n})\big)$$

More precisely, i.e. with an explicit O function, we obtain here from (6.4) and (6.1) and $U_1 = \lambda_1 = 2^{3n}$:

$$\lambda_\alpha \leq U_\alpha \Big(\frac{1 - \frac{3\alpha}{2^n} + \frac{3\alpha^2}{2^{2n}} - \frac{\alpha^3}{2^{3n}}}{1 - \frac{4\alpha}{2^n} + \frac{6\alpha^2}{2^{2n}} - \frac{4\alpha^3}{2^{3n}} + \frac{\alpha^4}{2^{4n}}} \Big)^\alpha$$

$$\lambda_\alpha \leq U_\alpha \Big(1 + \Big(\frac{\frac{\alpha}{2^n} - \frac{3\alpha^2}{2^{2n}} + \frac{3\alpha^3}{2^{3n}} - \frac{\alpha^4}{2^{4n}}}{1 - \frac{4\alpha}{2^n}} \Big)^\alpha \Big)$$

$$\lambda_\alpha \leq U_\alpha \Big(1 + \frac{\alpha^2}{2^n (1 - \frac{4\alpha}{2^n})} \Big)$$

since $\alpha \leq 2^n$.

Similarly, $\lambda_\alpha \geq U_\alpha \Big(1 - \frac{\alpha^3}{2^n (1 - \frac{4\alpha}{2^n})} \Big)$.

$$U_\alpha (1 - \frac{\alpha^3}{2^{2n} (1 - \frac{4\alpha}{2^n})}) \leq \lambda_\alpha \leq U_\alpha \Big(1 + \frac{\alpha^2}{2^n (1 - \frac{4\alpha}{2^n})} \Big) \quad (6.5) \quad (\underline{\text{“First Approximation”}})$$

i.e. from $\lambda_\alpha \leq U_\alpha (1 + O(\frac{\alpha^2}{2^n}))$, we get security until $\alpha^2 \ll 2^n$, i.e. until $\alpha \ll \sqrt{2^n}$. However, we want security until $\alpha \ll 2^n$ and not only $\alpha \ll \sqrt{2^n}$, so we want a better evaluation for $\frac{\lambda_{\alpha+1}}{2^{3n} \cdot \lambda_\alpha}$ (i.e. we want something like (6.3) instead of (6.4)).

7 An induction formula on λ_α ("Orange Equations")

A more precise evaluation

For each i, $1 \le i \le 4\alpha$, we will denote by B_i the set of $(f_1, \ldots, f_{\alpha+1}, g_1, \ldots, g_{\alpha+1}, h_1, \ldots, h_{\alpha+1})$, that satisfy the conditions λ_α and the conditions β_i. Therefore we have:

$$\lambda_{\alpha+1} = 2^{3n}\lambda_\alpha - |\cup_{i=1}^{4\alpha} B_i|$$

We know that for any set A_i and any integer μ, we have:

$$|\cup_{i=1}^{\mu} A_i| = \sum_{i=1}^{\mu} |A_i| - \sum_{i_1<i_2} |A_{i_1} \cap A_{i_2}| + \sum_{i_1<i_2<i_3} |A_{i_1} \cap A_{i_2} \cap A_{i_3}| + \ldots + (-1)^{\mu+1}|A_1 \cap A_2 \cap \ldots \cap A_\mu|$$

Moreover, each set of 5 (or more) equations β_i is in contradiction with the conditions λ_α because we will have at least two equations in f, or two in g, or two in h, or two in $f \oplus g \oplus h$ (and $f_{\alpha+1} = f_i$ and $f_{\alpha+1} = f_j$ gives $f_i = f_j$ with $i \ne j$ and $1 \le \alpha, j \le \alpha$, in contradiction with λ_α).

Therefore, we have:

$$\lambda_{\alpha+1} = 2^{3n}\lambda_\alpha - \sum_{i=1}^{4\alpha} |B_i| + \sum_{i<j} |B_i \cap B_j| - \sum_{i<j<k} |B_i \cap B_j \cap B_k| + \sum_{i<j<k<l} |B_i \cap B_j \cap B_k \cap B_l|$$

• 1 **equation**.

In B_i, we have the conditions λ_α plus the equation β_i, and β_i will fix $f_{\alpha+1}$, or $g_{\alpha+1}$, or $h_{\alpha+1}$ from the other values. Therefore:

$$|B_i| = 2^{2n}\lambda_\alpha \quad \text{and} \quad -\sum_{i=1}^{4\alpha} |B_i| = -4\alpha \cdot 2^{2n}\lambda_\alpha$$

• 2 **equations**.

First Case: β_i and β_j are two equations in f (or two in g, or two in h, or two in $f \oplus g \oplus h$. (For example: $f_{\alpha+1} = f_1$ and $f_{\alpha+2} = f_2$). Then these equations are not compatible with the conditions λ_α, therefore $|B_i \cap B_j| = 0$.

Second Case: we are not in the first case. Then two variables (for example f_α and g_α) are fixed from the others. Therefore: $|B_i \cap B_j| = 2^n\lambda_\alpha$ and $\sum_{i<j} |B_i \cap B_j| = 6\alpha^2 \cdot 2^n\lambda_\alpha$. ($6 = \binom{4}{2}$ is here the choice of 2 variables between f, g, h and $f \oplus g \oplus h$).

• 3 **equations**.

If we have two equations in f, or in g, or in h, or in $f \oplus g \oplus h$, we have $|B_i \cap B_j \cap B_k| = 0$. If we are not in these cases, then $f_{\alpha+1}$, $g_{\alpha+1}$ and $h_{\alpha+1}$ are fixed by the three equations from the other variables, and then $|B_i \cap B_j \cap B_k| = \lambda_\alpha$. Therefore: $-\sum_{i<j<k} |B_i \cap B_j \cap B_k| = -4\alpha^3\lambda_\alpha$. (4 comes from the fact we do not have an equation in f, g, h or in $f \oplus g \oplus h$).

• 4 **equations**.

This value is different from 0 only if we have one equation $f_{\alpha+1} = f_i$, one equation $g_{\alpha+1} = g_j$, one equation $h_{\alpha+1} = h_k$ and one equation $f_{\alpha+1} \oplus g_{\alpha+1} \oplus h_{\alpha+1} = f_l \oplus g_l \oplus h_l$. Then $|B_i \cap B_j \cap B_k \cap B_l|$ = number of f_a, g_b, h_c, with $a, b, c \in \{1, \ldots, \alpha\}$, that satisfy the conditions λ_α plus the equation X: $f_i \oplus g_j \oplus h_k = f_l \oplus g_l \oplus h_l$. We will denote by $\lambda'_\alpha(X)$ this number, and by λ'_α any value $\lambda'_\alpha(X)$ when X is linearly independent with the 4α conditions β_i.

Case 1. i, j, k, l are pairwise distinct. Here we have $\alpha(\alpha-1)(\alpha-2)(\alpha-3) = \alpha^4 - 6\alpha^3 + 11\alpha^2 - 6\alpha$ possibilities for i, j, k, l and from the symmetries of all indexes in the conditions λ_α, all the $\lambda'_\alpha(X)$ of this case 1 are equal. We denote by $\lambda_\alpha^{'(4)}$ this value of $\lambda'_\alpha(X)$. (The (4) here is to remember that we have exactly 4 indexes i, j, k, l). Typical equation X: $f_1 \oplus g_2 \oplus h_3 = f_4 \oplus g_4 \oplus h_4$.

Case 2. In $\{i, j, k, l\}$, we have exactly 3 indexes. Here we have $6\alpha(\alpha-1)(\alpha-2) = 6\alpha^3 - 18\alpha^2 + 12\alpha$ possibilities for i, j, k, l (since there are 6 possibilities to choose an equality). From the symmetries in the conditions λ_α, all the $\lambda'_\alpha(X)$ of this case 2 are equal. We denote by $\lambda_\alpha^{'(3)}$ this value of $\lambda'_\alpha(X)$. Typical equation X: $f_1 \oplus g_1 = f_2 \oplus g_3$ or $f_1 \oplus g_1 \oplus h_2 = f_3 \oplus g_3 \oplus h_3$.

Case 3. In $\{i, j, k, l\}$, 3 indexes have the same value (example $i = j = k$) and the other one has a different value. Then X is not compatible with the conditions λ_α.

Case 4. In i, j, k, l, we have 2 indexes and we are not in the Case 3 (for example $i = j$ and $k = l$). Here we have $3\alpha(\alpha-1) = 3\alpha^2 - 3\alpha$ possibilities for i, j, k, l. From the symmetries in the conditions λ_α all the $\lambda'_\alpha(X)$ of this case 4 are equal. We denote by $\lambda_\alpha^{'(2)}$ this value of $\lambda'_\alpha(X)$. Typical equation X: $f_1 \oplus g_1 = f_2 \oplus g_2$.

Case 5. We have $i = j = k = l$. Here we have α possibilities for i, j, k, l. Here X is always true, and $\lambda'_\alpha(X) = \lambda_\alpha$.

From these 5 cases we get:

$$\sum_{i<j<k<l} |B_i \cap B_j \cap B_k \cap B_l| = \alpha(\alpha-1)(\alpha-2)(\alpha-3)\lambda_\alpha^{'(4)} + 6\alpha(\alpha-1)(\alpha-2)\lambda_\alpha^{'(3)} + 3\alpha(\alpha-1)\lambda_\alpha^{'(2)} + \alpha\lambda_\alpha$$

Therefore (Exact "Orange Equations"):

$$\underline{\lambda_{\alpha+1} = (2^{3n} - 4\alpha \cdot 2^{2n} + 6\alpha^2 \cdot 2^n - 4\alpha^3 + \alpha)\lambda_\alpha + (\alpha^4 - 6\alpha^3 + 11\alpha^2 - 6\alpha)\lambda_\alpha^{'(4)} +}$$

$$\underline{(6\alpha^3 - 18\alpha^2 + 12\alpha)\lambda_\alpha^{'(3)} + (3\alpha^2 - 3\alpha)\lambda_\alpha^{'(2)}} \quad (7.1)$$

As said above, we denote by λ'_α any value of $\lambda'_\alpha(X)$ such that X is linearly independent with the 4α conditions β_i. Then, from (7.1) we write ("Orange Equations"):

$$\underline{\lambda_{\alpha+1} = (2^{3n} - 4\alpha \cdot 2^{2n} + 6\alpha^2 \cdot 2^n - 4\alpha^3 + \alpha)\lambda_\alpha + (\alpha^4 - 4\alpha^2 + 3\alpha)\lambda'_\alpha} \quad (7.2)$$

where $A \cdot \lambda'_\alpha$ is just a notation to mean that we have A terms λ'_α but each of these λ'_α may have different values. It is interesting to compare (6.1) on $U_{\alpha+1}$ with (7.2) on $\lambda_{\alpha+1}$. Our aim is to get (6.3) from (7.2). For this we see that we have to prove that

$$\lambda'_\alpha = \frac{\lambda_\alpha}{2^n}\left(1 + O\left(\frac{1}{2^n}\right) + O\left(\frac{\alpha}{2^{2n}}\right)\right) \quad (7.3)$$

for "most" values λ'_α or for the values $\lambda_\alpha^{'(4)}$. This is what we will do.

Remark. In fact, in (7.3), we only need

$$\lambda'_\alpha \leq \frac{\lambda_\alpha}{2^n}\left(1 + O\left(\frac{1}{2^n}\right) + O\left(\frac{\alpha}{2^{2n}}\right)\right)$$

for our results.

Strong $[\lambda'_\alpha]$

(The concept of "Strong $[\lambda'_\alpha]$" is not needed for our main results, but it will slightly improve some coefficient if we use it).

Definition 1 *We will say that an equation X is "strong", when X is not the Xor of a constant and of one or two equations of this type:*

$$f_i = f_j,\ g_i = g_j,\ h_i = h_j,\ \text{or}\ f_i \oplus g_i \oplus h_i = f_j \oplus g_j \oplus h_j$$

Similarly we will say that a coefficient λ'_α is "strong", and we denote it by $\lambda^{*'}_\alpha$ when the equation X of λ'_α is strong.

For example here, $\lambda_\alpha^{'(4)}$ (with typical $X : f_1 \oplus g_2 \oplus h_3 = f_4 \oplus g_4 \oplus h_4$) is "strong", but $\lambda_\alpha^{'(3)}$ (with typical $X : f_1 \oplus g_1 = f_2 \oplus g_3$ or $f_1 \oplus g_1 \oplus h_2 = f_3 \oplus g_3 \oplus h_3$) and $\lambda_\alpha^{'(2)}$ (with typical $X : f_1 \oplus g_1 = f_2 \oplus g_2$) are not strong since when $f_1 = f_2$, from $f_1 \oplus g_1 = f_2 \oplus g_3$, we get $g_1 = g_3$.
Therefore we can write ("Orange Equations" with strong λ'_α):

$$\underline{\lambda_{\alpha+1} = (2^{3n} - 4\alpha \cdot 2^{2n} + 6\alpha^2 \cdot 2^n - 4\alpha^3 + \alpha)\lambda_\alpha}$$

$$\underline{+(\alpha^4 - 6\alpha^3 + 11\alpha^2 - 6\alpha)\lambda^{*'}_\alpha + (6\alpha^3 - 15\alpha^2 + 9\alpha)\lambda'_\alpha} \quad (7.5)$$

8 Evaluations of $[\lambda'_\alpha]/\lambda_\alpha$ in $O(\frac{\alpha}{2^n})$, Security in $m \ll 2^{\frac{5n}{6}}$

By definition $[\lambda'_\alpha]$ denotes (as we have seen in Section 7) the number of

$$(f_1, \ldots, f_\alpha, g_1, \ldots, g_\alpha, h_1, \ldots, h_\alpha) \text{ of } I_n^{3\alpha}$$

that satisfy the conditions λ_α plus an equation X of the type:

$$f_j \oplus g_j \oplus h_j = f_k \oplus g_l \oplus h_i$$

with $i, j, k, l \in \{1, \ldots, \alpha\}$ such that X is compatible with the conditions λ_α and such that X is not $0 = 0$ (i.e. we do not have $i = j = k = l$). We have seen in Section 7 that $[\lambda'_\alpha]$ is not a fixed value: it can be $\lambda_\alpha^{'(4)}$ (by symmetries of the hypothesis for this case we can assume X to be: $f_\alpha \oplus g_\alpha \oplus h_\alpha = h_{\alpha-1} \oplus g_{\alpha-2} \oplus f_{\alpha-3}$) or $\lambda_\alpha^{'(3)}$ (for this case we can assume X to be: $f_\alpha \oplus g_\alpha = f_{\alpha-1} \oplus g_{\alpha-2}$) or $\lambda_\alpha^{'(2)}$ (for this case we can assume X to be: $f_\alpha \oplus g_\alpha = f_{\alpha-1} \oplus g_{\alpha-1}$). However, as we will see all these three values $[\lambda'_\alpha]$ are very near, and they are very near $\frac{\lambda_\alpha}{2^n}$. (Remark: we are mainly interested in $\lambda_\alpha^{'(4)}$ very near $\frac{\lambda_\alpha}{2^n}$ since in formula (7.1) of Section 7 we have a term in $\alpha^4 \lambda_\alpha^{'(4)}$).

Theorem 3 *For all values $[\lambda'_\alpha]$ we have:*

$$1 - \frac{8\alpha}{2^n} \leq \frac{2^n\ [\lambda'_\alpha]}{\lambda_\alpha} \leq 1 + \frac{8\alpha}{(1 - \frac{8\alpha}{2^n})2^n}$$

Proof of Theorem 3

We will present here the proof with $X : f_\alpha \oplus g_\alpha \oplus h_\alpha = h_{\alpha-1} \oplus g_{\alpha-2} \oplus f_{\alpha-3}$. The proof is exactly similar for all the other cases. From (6.4), we have:

$$1 - \frac{4(\alpha - 1)}{2^n} \leq \frac{\lambda_\alpha}{2^{3n}\lambda_{\alpha-1}} \leq 1$$

and

$$1 - \frac{4(\alpha-2)}{2^n} \leq \frac{\lambda_{\alpha-1}}{2^{3n}\lambda_{\alpha-2}} \leq 1$$

Therefore

$$2^{6n}\lambda_{\alpha-2}\big(1 - \frac{4(\alpha-1)}{2^n}\big)^2 \leq \lambda_\alpha \leq 2^{6n}\lambda_{\alpha-2} \quad (B1)$$

We will now evaluate $[\lambda'_\alpha]$ from $\lambda_{\alpha-2}$.
Remark: we evaluate here from $\lambda_{\alpha-2}$ and not from $\lambda_{\alpha-1}$ in order to have a variable $h_{\alpha-1}$ not fixed when we will combine the conditions 8 and 9 below.

In $[\lambda'_\alpha]$ we have the condition $\lambda_{\alpha-2}$ plus

1. $f_{\alpha-1} \notin \{f_1, \dots, f_{\alpha-2}\}$
2. $g_{\alpha-1} \notin \{g_1, \dots, g_{\alpha-2}\}$
3. $h_{\alpha-1} \notin \{h_1, \dots, h_{\alpha-2}\}$
4. $f_{\alpha-1} \oplus g_{\alpha-1} \oplus h_{\alpha-1} \notin \{f_1 \oplus g_1 \oplus h_1, \dots, f_{\alpha-2} \oplus g_{\alpha-2} \oplus h_{\alpha-2}\}$
5. $f_\alpha \notin \{f_1, \dots, f_{\alpha-1}\}$
6. $g_\alpha \notin \{g_1, \dots, g_{\alpha-1}\}$
7. $h_\alpha \notin \{h_1, \dots, h_{\alpha-1}\}$
8. $f_\alpha \oplus g_\alpha \oplus h_\alpha \notin \{f_1 \oplus g_1 \oplus h_1, \dots, f_{\alpha-1} \oplus g_{\alpha-1} \oplus h_{\alpha-1}\}$
9. (Equation X): $f_\alpha \oplus g_\alpha \oplus h_\alpha = f_{\alpha-3} \oplus g_{\alpha-2} \oplus h_{\alpha-1}$

We can decide that X will fix h_α from the other values: $h_\alpha = f_\alpha \oplus g_\alpha \oplus f_{\alpha-3} \oplus g_{\alpha-2} \oplus h_{\alpha-1}$, and we can decide that conditions 2., 3., 4. and 8. will be written in $h_{\alpha-1}$ and $g_{\alpha-1}$:

$$\begin{gathered} h_{\alpha-1} \notin \big\{h_1, \dots, h_{\alpha-2}, \\ f_1 \oplus g_1 \oplus h_1 \oplus f_{\alpha-1} \oplus g_{\alpha-1}, \dots, f_{\alpha-2} \oplus g_{\alpha-2} \oplus h_{\alpha-2} \oplus f_{\alpha-1} \oplus g_{\alpha-1}, \\ f_1 \oplus g_1 \oplus h_1 \oplus f_{\alpha-3} \oplus g_{\alpha-2}, \dots, f_{\alpha-2} \oplus h_{\alpha-2} \oplus f_{\alpha-3}\big\} \end{gathered}$$

In this set we have between $\alpha-2$ and $3(\alpha-2)$ elements when $h_1, \dots, h_{\alpha-2}$ are pairwise distinct.

$$g_{\alpha-1} \notin \big\{g_1, \dots, g_{\alpha-2},\ f_{\alpha-1} \oplus f_{\alpha-3} \oplus g_{\alpha-2}\big\}$$

In this set we have between $\alpha-2$ and $\alpha-1$ elements when $g_1, \dots, g_{\alpha-2}$ are pairwise distinct ($g_{\alpha-1} \neq f_{\alpha-1} \oplus f_{\alpha-3} \oplus g_{\alpha-2}$ comes from the last condition 8).

Similarly, we can write conditions 6 and 7 in g_α:

$$g_\alpha \notin \big\{g_1, \dots, g_{\alpha-1},\ h_1 \oplus f_\alpha \oplus f_{\alpha-3} \oplus g_{\alpha-2} \oplus h_{\alpha-1}, \dots, h_{\alpha-1} \oplus f_\alpha \oplus f_{\alpha-3} \oplus g_{\alpha-2} \oplus h_{\alpha-1}\big\}$$

In this set we have between $\alpha-1$ and $2(\alpha-1)$ elements when $g_1, \dots, g_{\alpha-1}$ are pairwise distinct. Therefore we get:

$$[\lambda'_\alpha] \geq \lambda_{\alpha-2} \underbrace{\big(2^n - (\alpha-2)\big)}_{f_{\alpha-1}} \underbrace{\big(2^n - (\alpha-1)\big)}_{g_{\alpha-1}} \underbrace{\big(2^n - 3(\alpha-2)\big)}_{h_{\alpha-1}} \underbrace{\big(2^n - (\alpha-1)\big)}_{f_\alpha} \underbrace{\big(2^n - 2(\alpha-1)\big)}_{g_\alpha}$$

and

$$[\lambda'_\alpha] \le \lambda_{\alpha-2} \underbrace{(2^n-(\alpha-2))}_{f_{\alpha-1}} \underbrace{(2^n-(\alpha-2))}_{g_{\alpha-1}} \underbrace{(2^n-(\alpha-2))}_{h_{\alpha-1}} \underbrace{(2^n-(\alpha-1))}_{f_\alpha} \underbrace{(2^n-(\alpha-1))}_{g_\alpha}$$

So

$$\big(1-\frac{(\alpha-2)}{2^n}\big)\big(1-\frac{(\alpha-1)}{2^n}\big)^2\big(1-\frac{3(\alpha-2)}{2^n}\big)\big(1-\frac{2(\alpha-1)}{2^n}\big) \le \frac{[\lambda'_\alpha]}{2^{5n}\lambda_{\alpha-2}} \le \big(1-\frac{(\alpha-2)}{2^n}\big)^3\big(1-\frac{(\alpha-1)}{2^n}\big)^2$$

So we have:

$$1-\frac{8\alpha}{2^n} \le \frac{[\lambda'_\alpha]}{2^{5n}\lambda_{\alpha-2}} \le 1$$

and with $(B1)$ this gives:

$$\frac{2^{5n}\lambda_\alpha}{2^{6n}}\big(1-\frac{8\alpha}{2^n}\big) \le [\lambda'_\alpha] \le \frac{2^{5n}\lambda_\alpha}{2^{6n}(1-\frac{4(\alpha-1)}{2^n})^2} \le \frac{\lambda_\alpha}{2^n(1-\frac{8\alpha}{2^n})}$$

So

$$1-\frac{8\alpha}{2^n} \le \frac{2^n[\lambda'_\alpha]}{\lambda_\alpha} \le 1+\frac{8\alpha}{2^n(1-\frac{8\alpha}{2^n})}$$

as claimed.

Theorem 4 *We have* $\mathrm{Adv}_\phi^{\mathrm{PRF}} \le 2\big(\frac{1}{E(h)}+\frac{m^2}{2^{3n}(1-\frac{4m}{2^n})}+\frac{8m^6}{2^{5n}(1-\frac{12m}{2^n})}\big)^{1/3} \quad (B.2)$

Proof of Theorem 4

This proof follows immediately from Theorem 3 and formula (8.6) of Section 8.

Remark: If $m >> \sqrt{2^n}$ (these are the only difficult cases), then in this expression, the main term is $\Big(\frac{8m^6}{2^{5n}(1-\frac{12m}{2^n})}\Big)^{1/3}$ in $O(\frac{m^2}{2^{5n/3}})$, i.e. we have security when $m \ll 2^{\frac{5n}{6}}$ ("Second Approximation").

In order to get security in $m \ll 2^n$, instead of $m \ll 2^{5n/6}$, we need to have a better evaluation of $[\lambda'_\alpha]$ (i.e. we need $|[\epsilon_\alpha]| = O(\frac{\alpha}{2^{2n}})$ instead of $O(\frac{\alpha}{2^n})$).

Part III
Security in $m \ll 2^{\frac{9n}{10}}$

9 From $[\epsilon_\alpha]$ to $\mathrm{Adv}_\phi^{\mathrm{PRF}}$

Let $[\epsilon_\alpha] = \frac{2^n[\lambda'_\alpha]}{\lambda_\alpha} - 1$. Therefore, $[\lambda'_\alpha] = \frac{\lambda_\alpha}{2^n}(1+[\epsilon_\alpha])$ and (7.3) means: $[\epsilon_\alpha] = O(\frac{1}{2^n}) + O(\frac{\alpha}{2^{2n}})$. From the analysis of the previous sections, we know that if we can prove that $|[\epsilon_\alpha]|$ is small, then $\mathrm{Adv}_\phi^{\mathrm{PRF}}$ will be small. Moreover, since we have done exact analysis, we know reciprocally that if Adv_ϕ^{PRF} is small, then

$|[\epsilon_\alpha]|$ must be small for $\lambda_\alpha^{'(4)}$. Let evaluate more precisely the links between $|[\epsilon_\alpha]|$ and $\mathrm{Adv}_\phi^{\mathrm{PRF}}$ that we have. From formula (7.2), we have:

$$\lambda_{\alpha+1} = 2^{3n}\big[1 - \frac{4\alpha}{2^n} + \frac{6\alpha^2}{2^{2n}} - \frac{4\alpha^3}{2^{3n}} + \frac{\alpha}{2^{3n}} + \frac{(\alpha^4 - 4\alpha^2 + 3\alpha)}{2^{4n}} + A\big]\lambda_\alpha$$

with

$$A \leq \frac{\alpha^4[\epsilon_\alpha]}{2^n \cdot 2^{3n}} \quad (8.1)$$

Therefore, by using U_α of section 6 we have:

$$\frac{\lambda_{\alpha+1}}{\lambda_\alpha} = \frac{U_{\alpha+1}}{U_\alpha} \cdot \frac{(1 - \frac{4\alpha}{2^n} + \frac{6\alpha^2}{2^{2n}} - \frac{4\alpha^3}{2^{3n}} + \frac{\alpha}{2^{3n}} + \frac{(\alpha^4 - 4\alpha^2 + 3\alpha)}{2^{4n}} + A)}{(1 - \frac{4\alpha}{2^n} + \frac{6\alpha^2}{2^{2n}} - \frac{4\alpha^3}{2^{3n}} + \frac{\alpha^4}{2^{4n}})}$$

$$\frac{\lambda_{\alpha+1}}{\lambda_\alpha} = \frac{U_{\alpha+1}}{U_\alpha} \cdot \big(1 + \frac{\frac{\alpha}{2^{3n}} - \frac{4\alpha^2}{2^{4n}} + \frac{3\alpha}{2^{4n}} + A}{1 - \frac{4\alpha}{2^n} + \frac{6\alpha^2}{2^{2n}} - \frac{4\alpha^3}{2^{3n}} + \frac{\alpha^4}{2^{4n}}}\big) \quad (8.2)$$

Therefore, with (8.1) we have

$$\frac{\lambda_{\alpha+1}}{\lambda_\alpha} = \frac{U_{\alpha+1}}{U_\alpha} \cdot \big(1 + O_1(\frac{\alpha}{2^{3n}}) + O_2(A)\big)$$

with

$$|O_1(\frac{\alpha}{2^{3n}})| \leq \frac{\alpha}{2^{3n}(1 - \frac{4\alpha}{2^n})} \quad (8.3)$$

and

$$|O_2(A)| \leq \frac{A}{(1 - \frac{4\alpha}{2^n})} \quad (8.4)$$

Since $\lambda_1 = U_1 = 2^{3n}$, we have

$$\lambda_\alpha = (\frac{\lambda_\alpha}{\lambda_{\alpha-1}})(\frac{\lambda_{\alpha-1}}{\lambda_{\alpha-2}}) \dots (\frac{\lambda_2}{\lambda_1})\lambda_1 = U_\alpha\big[1 + O(\frac{\alpha}{2^{3n}}) + O(A)\big]^\alpha$$

$$\lambda_\alpha = \frac{[2^n(2^n-1)\dots(2^n-\alpha+1)]^4}{2^{n\alpha}}\big(1 + O(\frac{\alpha^2}{2^{3n}}) + \alpha O(A)\big) \quad (8.5)$$

Now from (8.5) and (5.5) we get:

$$V(N) \leq E(N) + (E(N))^2\big(O(\frac{\alpha^2}{2^{3n}}) + \alpha O(A)\big)$$

Therefore, from (3.1) we get that the best CPA-2 attacks ϕ satisfy:

$$\mathrm{Adv}_\phi^{PRF} \leq 2\Big(\frac{V(N)}{E^2(N)}\Big)^{1/3} \leq 2\Big(\frac{1}{E(N)} + O(\frac{\alpha^2}{2^{3n}}) + \alpha O(A)\Big)^{1/3}$$

More precisely, by using (8.3) and (8.4) we get:

$$\mathrm{Adv}_\phi^{PRF} \leq 2\Big(\frac{1}{E(N)} + \frac{m^2}{2^{3n}(1 - \frac{4m}{2^n})} + \frac{\alpha^5 \cdot [\epsilon_\alpha]}{2^{4n} \cdot (1 - \frac{4\alpha}{2^n})}\Big)^{1/3} \quad (8.6)$$

Here we have $\frac{1}{E(N)} = \frac{2^{nm}}{(2^n(2^n-1)...(2^n-m+1))^2}$ and this is much smaller than $\frac{m^3}{2^{3n}}$ for example, thanks to Stirling Formula. From formula (8.6) we see clearly that a bound on $|[\epsilon_\alpha]|$ gives immediately a precise bound on Adv_ϕ^{PRF}. Now, we will present good bounds for $|[\epsilon_\alpha]|$. We will proceed progressively: first, we have obtained a bound for $|[\epsilon_\alpha]|$ in $O(\frac{\alpha}{2^n})$ and therefore a security (from (8.6)) in $O(2^{\frac{5n}{6}})$. Then we will get a bound for $|[\epsilon_\alpha]|$ in $O(\frac{\alpha^5}{2^{5n}})$ and therefore a security (from (8.6)) in $O(2^{\frac{9n}{10}})$. Finally, we will iterate the process in order to obtain security in $m \ll O(2^n)$ as wanted.

10 Security in $m \ll 2^{\frac{9n}{10}}$

We will denote by $[\epsilon_\alpha] = \frac{2^n[\lambda'_\alpha]}{\lambda_\alpha} - 1$. Therefore, $[\lambda'_\alpha] = \frac{\lambda_\alpha}{2^n}(1+[\epsilon_\alpha])$. We have seen in Appendix 8 (Theorem 3) that $|[\epsilon_\alpha]| \le \frac{8\alpha}{2^n\cdot(1-\frac{8\alpha}{2^n})}$. We want now a better evaluation of $|[\epsilon_\alpha]|$, since this will give us (cf formula (8.6)) a better security result. If we write formula $(C.1)$ of Appendix C in $[\epsilon_\alpha]$ instead of $[\lambda'_\alpha]$, we get:

$$2^n\lambda'^{(4)}_{\alpha+1} = \big[2^{3n} - 4\alpha\cdot 2^{2n} + (6\alpha^2-1)\cdot 2^n + (-4\alpha^3+5\alpha-3)\big]\lambda_\alpha +$$

$$(-\alpha\cdot 2^{2n} + 3\alpha^2\cdot 2^n - 4\alpha^3 + 5\alpha - 3)[\epsilon_\alpha]\lambda_\alpha + (\alpha^4 - 7\alpha^2 + 5\alpha)\cdot 2^n[\lambda''_\alpha] \quad (D.1)$$

Similarly, if we write formula (7.2) of Section 7 in $[\epsilon_\alpha]$ instead of $[\lambda'_\alpha]$, we get:

$$\lambda_{\alpha+1} = \big(2^{3n} - 4\alpha\cdot 2^{2n} + 6\alpha^2\cdot 2^n - 4\alpha^3 + \alpha + \frac{\alpha^4-4\alpha^2+3\alpha}{2^n}\big)\lambda_\alpha + \frac{(\alpha^4-4\alpha^2+3\alpha)}{2^n}[\epsilon_\alpha]\lambda_\alpha \quad (D.2)$$

Therefore, for $[\epsilon_{\alpha+1}] = \frac{2^n\lambda'^{(4)}_{\alpha+1} - \lambda_{\alpha+1}}{\lambda_{\alpha+1}}$, we obtain:

$$[\epsilon_{\alpha+1}] = (-2^n + (4\alpha-3) + \frac{-\alpha^4+4\alpha^2-3\alpha}{2^n})\frac{\lambda_\alpha}{\lambda_{\alpha+1}} +$$

$$(-\alpha\cdot 2^{2n} + 3\alpha^2\cdot 2^n - 4\alpha^3 + 5\alpha - 3 + \frac{-\alpha^4+4\alpha^2-3\alpha}{2^n})\frac{[\epsilon_\alpha]\lambda_\alpha}{\lambda_{\alpha+1}} +$$

$$(\alpha^4 - 7\alpha^2 + 5\alpha)2^n\frac{[\lambda''_\alpha]}{\lambda_{\alpha+1}} \quad (D.3)$$

Therefore:

$$[\epsilon_{\alpha+1}] = -[\epsilon_\alpha]\frac{\alpha}{2^n} + \frac{\alpha^4}{2^n\cdot\lambda_{\alpha+1}}(2^{2n}[\lambda''_\alpha] - \lambda_\alpha) + \mathrm{negl} \quad (D.4)$$

Where "negl" are some terms negligible compared with $[\epsilon_\alpha]\frac{\alpha}{2^n}$, or negligible compared with $O(\frac{1}{2^n})$. Now, exactly as we have proved (cf section 8):

$$1 - \frac{8\alpha}{2^n} \le \frac{2^n[\lambda'_\alpha]}{\lambda_\alpha} \le 1 + \frac{8\alpha}{(1-\frac{8\alpha}{2^n})2^n}$$

we have

$$\big(1 - \frac{8\alpha}{2^n}\big)^2 \le \frac{2^{2n}[\lambda''_\alpha]}{\lambda_\alpha} \le \frac{1}{(1-\frac{8\alpha}{2^n})^2} \quad (D.5)$$

(Evaluation in $O(\frac{\alpha}{2^n})$ are easy, we just have to proceed like in Appendix B. Evaluation in $O(\frac{1}{2^n})$ or in $O(\frac{\alpha}{2^{2n}})$ are more difficult). From $(D.4)$ we have:

$$|[\epsilon_{\alpha+1}]| \leq [\epsilon_\alpha]\frac{\alpha}{2^n} + \frac{\alpha^4}{2^n}\Big(\frac{2^{2n}[\lambda''_\alpha] - \lambda_\alpha}{\lambda_\alpha}\Big)\Big(\frac{\lambda_\alpha}{\lambda_{\alpha+1}}\Big) + \text{negl} \quad (D.6)$$

Now from (6.4) we have: $\frac{\lambda_\alpha}{\lambda_{\alpha+1}} \leq \frac{1}{(1-\frac{4\alpha}{2^n})\cdot 2^{3n}}$. Therefore, from $(D.6)$, (6.4) and $(D.5)$ we get:

$$|[\epsilon_{\alpha+1}]| \leq [\epsilon_\alpha]\frac{\alpha}{2^n} + \frac{\alpha^5}{2^{5n}} \cdot \frac{1}{(1-\frac{20\alpha}{2^n})} + \text{negl} \quad (D.7)$$

Now from our first approximation $|[\epsilon_\alpha]| \leq \frac{8\alpha}{(1-\frac{8\alpha}{2^n})2^n}$, we get from $(D.7)$

$$|[\epsilon_{\alpha+1}]| \leq \frac{8\alpha^2}{2^{2n}} + o(\frac{\alpha^2}{2^{2n}})$$

(where $o(\frac{\alpha^2}{2^{2n}})$ are negligible terms compared with $\frac{\alpha^2}{2^{2n}}$). By re-injecting this in $(D.7)$ we get:

$$|[\epsilon_{\alpha+2}]| \leq \frac{8\alpha^3}{2^{3n}} + o(\frac{\alpha^3}{2^{3n}})$$

One more time

$$|[\epsilon_{\alpha+3}]| \leq \frac{8\alpha^4}{2^{4n}} + o(\frac{\alpha^4}{2^{4n}})$$

One more time

$$|[\epsilon_{\alpha+4}]| \leq \frac{9\alpha^5}{2^{5n}} + o(\frac{\alpha^5}{2^{5n}})$$

Therefore, if $\alpha \geq 5$, we can write:

$$\underline{|[\epsilon_\alpha]| \leq \frac{9\alpha^5}{2^{5n}} + o(\frac{\alpha^5}{2^{5n}})} \quad (D.8 : \text{second evaluation of } [\epsilon_\alpha])$$

Now from $(D.8)$ and (8.6) we get:

$$\text{Adv}_\phi^{\text{PRF}} \leq 2\Big(\frac{1}{E(N)} + \frac{m^2}{2^{3n}(1-\frac{4m}{2^n})} + \frac{9m^{10}}{2^{9n}(1-\frac{4m}{2^n})} + o(\frac{m^{10}}{2^{9n}})\Big)^{\frac{1}{3}} \quad (D.9)$$

Therefore here we have obtained security when $m \ll 2^{\frac{9n}{10}}$.

11 Security in $m \ll 2^{\frac{kn}{k+1}}$ for any integer k and security in $m \ll 2^n$

In Section 6, we have obtained security when $m \ll \sqrt{2^n}$. In Appendix B, we have obtained security when $m \ll 2^{\frac{5n}{6}}$. In Appendix D, we have obtained security when $m \ll 2^{\frac{9n}{10}}$. Moreover, what we did in Appendix D is just the same thing as in Appendix B, with the analysis of λ''_α values (with 2 more equations X and Y than λ_α) in a similar way of λ'_α values (with one more equation X than λ_α). Obviously, we can iterate the process by introducing λ'''_α (with 3 more equations) in the same way etc. With λ'''_α we will obtain a better evaluation for $\frac{[\lambda''_\alpha]}{\lambda_\alpha}$ and from it and formula $(D.4)$, it will give us a better bound for $|[\epsilon_\alpha]|$. If we look at the process of the proof that we use here (in order to obtain proofs of security in $2^{\frac{kn}{k+1}}$ for larger and larger k) we see that we use two types of relations:

1. Evaluation in $O(\frac{\alpha}{2n})$ of $\frac{\lambda'_\alpha}{\lambda_\alpha}$, $\frac{\lambda''_\alpha}{\lambda_\alpha}$, $\frac{\lambda'''_\alpha}{\lambda_\alpha}$ etc. This is easily obtained as in Appendix B. More precisely, by iterating the evaluation of Appendix B, we get:

$$\left(1-\frac{8\alpha}{2^n}\right)^\mu \leq \frac{2^{kn}[\lambda_\alpha^{[\mu]}]}{\lambda_\alpha} \leq \frac{1}{(1-\frac{8\alpha}{2^n})^\mu}$$

for any integer k. ($[\mu]$ means that we have μ more equations, compatible and independent, in $\lambda_\alpha^{[\mu]}$ than in λ_α).

2. We have an induction formula that gives $\lambda_{\alpha+1}^{[\mu]}$ from values $\lambda_i^{[\mu-1]}$, $\lambda_i^{[\mu]}$, $\lambda_i^{[\mu+1]}$, $i \leq \alpha$.

By combining 1 and 2, we get security better and better when μ increases. More precisely, if we look at the number of operations that we perform in order to obtain security in $m \ll 2^{\frac{kn}{k+1}}$, we see that the coefficient involved increases at most in 2^k. Therefore from (8.6) we will get:

$$\mathrm{Adv}_\phi^{\mathrm{PRF}} \leq 2\Big(\frac{1}{E(N)} + \frac{m^2}{2^{3n}(1-\frac{4m}{2^n})} + \sum_{i=1}^{k}\frac{(2m)^i}{2^{in}} + \frac{9m^{k+1}}{2^{kn}(1-\frac{4m}{2^n})}\Big)^{\frac{1}{3}}$$

(In formula $(D.9)$ the term $\sum_{i=1}^{9}\frac{(2m)^i}{2^{in}}$ was in the $o(\frac{m^{10}}{2^{9n}})$). Therefore,

$$\mathrm{Adv}_\phi^{\mathrm{PRF}} \leq 2\Big(\frac{1}{E(N)} + \frac{m^2}{2^{3n}(1-\frac{4m}{2^n})} + \frac{2m}{2^n(1-\frac{2m}{2^n})} + \frac{9m^{k+1}}{2^{kn}(1-\frac{4m}{2^n})}\Big)^{\frac{1}{3}} \quad (E.1)$$

This gives security when $m \ll 2^{\frac{kn}{k+1}}$ for any integer k. Finally, by choosing $k = n$, we can notice that $\frac{m^{n+1}}{2^{(n^2)}} \leq \left(\frac{2m}{2^n}\right)^n$ (since $m \leq 2^n$). Therefore we have:

$$\underline{\mathrm{Adv}_\phi^{\mathrm{PRF}} \leq 2\Big(\frac{1}{E(N)} + \frac{m^2}{2^{3n}(1-\frac{4m}{2^n})} + \frac{2m}{2^n(1-\frac{2m}{2^n})} + \left(\frac{2m}{2^n}\right)^n \cdot \frac{9}{(1-\frac{4m}{2^n})}\Big)^{\frac{1}{3}}} \quad (E.2)$$

This gives security when $m \ll 2^n$, as wanted.

Part IV
General Security results

12 The general induction formula

Notations
Let α and β be two integers. We write $[\lambda_\alpha^d]$, or simply λ_α^d for simplicity, the number of sequences (f_i, g_i, h_i), $1 \leq i \leq \alpha$, $f_i \in I_n$, $g_i \in I_n$, $h_i \in I_n$ such that:

1. The f_i are pairwise distinct, $1 \leq i \leq \alpha$.
2. The g_i are pairwise distinct, $1 \leq i \leq \alpha$.
3. The h_i are pairwise distinct, $1 \leq i \leq \alpha$.

4. The $f_i \oplus g_i \oplus h_i$ are pairwise distinct, $1 \le i \le \alpha$.
5. We have d independent and compatible affine equations $X_1, X_2, \dots, X_d$ in the variables f_i, g_i, h_i, $1 \le i \le \alpha$. Here by "compatible" we mean that by linearity from $X_1, X_2, \dots, X_d$, we cannot obtain an equation $f_i = f_j$, or $g_i = g_j$, or$h_i = h_j$, or $f_i \oplus g_i \oplus h_i = f_j \oplus g_j \oplus h_j$, with $i \neq j$.
6. We assume that all these equations X_k, $1 \le k \le d$ can be written like this:
f_k (or g_k or h_k or $f_k \oplus g_k \oplus h_k$) $= \oplus$ of terms of indices $\le k-1$ in $f_i, g_i, h_i \oplus \psi_k$, where ψ_k is a constant of I_n. (We need $\psi_k = 0$ for our final results, but it is sometimes useful in some proofs to obtain some results with $\psi_k \neq 0$ as well).

Remark.
λ_α^d is a simple notation for $\lambda_\alpha^d(X_1, X_2, \dots, X_d)$, i.e. the values λ_α^d generally depend on $X_1, X_2, \dots, X_d$. However, as we will see, all these values λ_α^d are often very near.
Notation: χ
We will denote by χ the number of indices i used in the d equations $X_1, X_2, \dots, X_d$ in the variables f_i, g_i, h_i.
Remark.
This value χ will help us to evaluate the number of new indices in new equations. Often in our systems we will have $\chi \ll \alpha$ (typically we can have $\alpha \ll 2^n$ and $\chi \ll n$).

We will now generalize formula $(C1)$of section 10, i.e. we will evaluate $\lambda_{\alpha+1}^{d+1}$ from λ_α^d, λ_α^{d+1} and λ_α^{d+2}.

Theorem 5 *("General induction formula")*
There are some real values ϵ_1, ϵ_2, ϵ_3, ϵ_4, ϵ_5, ϵ_6, such that $\forall i \in \{1,2,3,4,5,6\}$, $0 \le \epsilon_i \le 1$, and:

$$\begin{gathered}\lambda_{\alpha+1}^{d+1} = 2^{2n}\lambda_\alpha^d \\ -3\alpha \cdot 2^n \lambda_\alpha^d - 2^{2n}\alpha\lambda_\alpha^{d+1} + \epsilon_1 \cdot \chi \cdot 2^{2n}\lambda_\alpha^{d+1} \\ +3\alpha^2\lambda_\alpha^d + 3\alpha^2 \cdot 2^n\lambda_\alpha^{d+1} - \epsilon_2 \cdot 3\chi\alpha \cdot 2^n \lambda_\alpha^{d+1} \\ -(4\alpha^3 - \epsilon_3\chi^3)\lambda_\alpha^{d+1} - \epsilon_3\chi^3\lambda_\alpha^d + \epsilon_4(12\alpha\chi^2)\lambda_\alpha^{d+1} \\ +(\alpha^4 - \epsilon_5 \cdot \alpha(\chi^3+1))\lambda_\alpha^{d+2} \\ +\epsilon_5 \cdot \alpha \cdot (\chi^3+1)\lambda_\alpha^{d+1} - \epsilon_6(6\chi^2\alpha^2 + \alpha^3\chi + 4\alpha)\lambda_\alpha^{d+2}\end{gathered}$$

Proof of Theorem 5

Part V
Variants and Conclusion

13 A simple variant of the schemes with only one permutation

Instead of $G = f_1 \oplus f_2$, $f_1, f_2 \in_R B_n$, we can study $G'(x) = f(x\|0) \oplus f(x\|1)$, with $f \in_R B_n$ and $x \in I_{n-1}$. This variant was already introduced in [2] and it is for this that in [2] p.9 the security in $\frac{m}{2^n} + O(n)\left(\frac{m}{2^n}\right)^{3/2}$ is presented. In fact, from a theoretical point of view, this variant G' is very similar to G, and it is possible to prove that our analysis can be modified to obtain a similar proof of security for G'.

14 A simple property about the Xor of two permutations and a new conjecture

I have conjectured this property:

$$\forall f \in F_n,\ \text{if} \bigoplus_{x \in I_n} f(x) = 0,\ \text{then}\ \exists (g,h) \in B_n^2,\ \text{such that}\ f = g \oplus h.$$

Just one day after this paper was put on eprint, J.F. Dillon pointed to us that in fact this was proved in 1952 in [5]. We thank him a lot for this information. (This property was proved again independently in 1979 in [16]).

A new conjecture. However I conjecture a stronger property. Conjecture:

$$\forall f \in F_n,\ \text{if} \bigoplus_{x \in I_n} f(x) = 0,\ \text{then the number}\ H\ \text{of}\ (g,h) \in B_n^2,$$

$$\text{such that}\ f = g \oplus h\ \text{satisfies}\ H \geq \frac{|B_n|^2}{2^{n2^n}}.$$

Variant: I also conjecture that this property is true in any group, not only with Xor.

Remark: in this paper, I have proved weaker results involving m equations with $m \ll O(2^n)$ instead of all the 2^n equations. These weaker results were sufficient for the cryptographic security wanted.

15 Conclusion

The results in this paper improve our understanding of the PRF-security of the Xor of two random permutations. More precisely in this paper we have proved that the Adaptive Chosen Plaintext security for this problem is in $O(2^n)$, and we have obtained an explicit O function. These results belong to the field of finding security proofs for cryptographic designs above the "birthday bound". (In [1, 8, 13], some results "above the birthday bound" on completely different cryptographic designs are also given). Since building PRF from PRP has many practical applications,we believe that these results are of real interest both from a theoretical point of view and a practical point of view. Our proofs need a few pages, so are a bit hard to read, but the results obtained are very easy to use and the mathematics used are elementary (essentially combinatorial and induction arguments). Moreover, we have proved (in Section 5) that this cryptographic problem of security is directly related to a very simple to describe and purely combinatorial problem. We have obtained this transformation by using the "H_σ technique", i.e. combining the "coefficient H technique" of [12, 13] and a specific computation of the standard deviation of H. (In a way, from a cryptographic point of view, this is maybe the most important result, and all the analysis after Section 5 can be seen as combinatorial mathematics and not cryptography anymore). It is also interesting to notice that in our proof with have proceeded with "necessary and sufficient" conditions, i.e. that the H_σ property that we proved is exactly equivalent to the cryptographic property that we wanted. Moreover, as we have seen, less strong results of security are quickly obtained.

References

[1] William Aiello and Ramarathnam Venkatesan. Foiling Birthday Attacks in Length-Doubling Transformations - Benes: A Non-Reversible Alternative to Feistel. In Ueli M. Maurer, editor, *Advances in*

Cryptology – EUROCRYPT '96, volume 1070 of *Lecture Notes in Computer Science*, pages 307–320. Springer-Verlag, 1996.

[2] Mihir Bellare and Russell Impagliazzo. A Tool for Obtaining Tighter Security Analyses of Pseudorandom Function Based Constructions, with Applications to PRP to PRF Conversion. ePrint Archive 1999/024: Listing for 1999.

[3] Mihir Bellare, Ted Krovetz, and Phillip Rogaway. Luby-Rackoff Backwards: Increasing Security by Making Block Ciphers Non-invertible. In Kaisa Nyberg, editor, *Advances in cryptology – EUROCRYPT 1998*, volume 1403 of *Lecture Notes in Computer Science*, pages 266–280. Springer-Verlag, 1998.

[4] Chris Hall, David Wagner, John Kelsey, and Bruce Schneier. Building PRFs from PRPs. In Hugo Krawczyk, editor, *Advances in Cryptology – CRYPTO 1998*, volume 1462 of *Lecture Notes in Computer Science*, pages 370–389. Springer-Verlag, 1998.

[5] Marshall Hall Jr. A Combinatorial Problem on Abelian Groups. *Proceedings of the Americal Mathematical Society*, 3(4):584–587, 1952.

[6] Stefan Lucks. The Sum of PRPs Is a Secure PRF. In Bart Preneel, editor, *Advances in Cryptology – EUROCRYPT 2000*, volume 1807 of *Lecture Notes in Computer Science*, pages 470–487. Springer-Verlag, 2000.

[7] Avradip Mandal, Jacques Patarin, and Valérie Nachef. Indifferentiability beyond the Birthday Bound for the Xor of Two Public Random Permutations. In Guang Gong and Kishan Chand Gupta, editors, *Progress in Cryptology – INDOCRYPT 2010*, volume 6948 of *Lecture Notes in Computer Science*, pages 69–81. Springer-Verlag, 2010.

[8] Ueli Maurer and Krzysztof Pietrzak. The Security of Many-Round Luby-Rackoff Pseudo-Random Permutations. In Eli Biham, editor, *Advances in Cryptology – EUROCRYPT 2003*, volume 2656 of *Lecture Notes in Computer Science*, pages 544–561. Springer-Verlag, 2003.

[9] Jacques Patarin. Generic Attacks for the Xor of k Random Permutations. *Paper accepted at ACNS 2013*.

[10] Jacques Patarin. Introduction to Mirror Theory: Analysis of Systems of Linear Equalities and Linear Non Equalitites for Cryptography. *Cryptology ePrint archive: 2010/287: Listing for 2010*.

[11] Jacques Patarin. Security in $0(2^n)$ for the Xor of Two Random Permutations - Proof with the standard H technique. . *Available from the author*.

[12] Jacques Patarin. Etude de Générateurs de Permutations Basés sur les Schémas du DES. In *Ph. Thesis*. Inria, Domaine de Voluceau, France, 1991.

[13] Jacques Patarin. Luby-Rackoff: 7 Rounds are Enough for $2^{n(1-\epsilon)}$ Security. In Dan Boneh, editor, *Advances in Cryptology – CRYPTO 2003*, volume 2729 of *Lecture Notes in Computer Science*, pages 513–529. Springer-Verlag, 2003.

[14] Jacques Patarin. On linear systems of equations with distinct variables and Small block size. In Dongho Wan and Seungjoo Kim, editors, *ICISC 2005*, volume 3935 of *Lecture Notes in Computer Science*, pages 299–321. Springer-Verlag, 2006.

Table 1: Summary of the results on λ_m for $m = 1, 2, 3$

$\lambda_1 = 2^{3n}$	$\lambda_2 = 2^{3n}(2^n - 1)(2^{2n} - 3.2^n + 3)$	$\lambda_3 = 2^{3n}.(2^n - 1)(2^n - 2)$ $.(2^{4n} - 9.3^{3n} + 33.2^{2n} - 60.2^n + 48)$
$\downarrow$ $Adv_1 = 0$	$\lambda_2^{'(2)} =$ $\lambda_2^{'(2)} = \lambda_2' = 2^{3n}.(2^n - 1)^2$	$\lambda_3^{'(3)} = 2^{3n}(2^n - 1)(2^n - 2)(2^n - 3)$ $.(2^{2n} - 5.2^n + 8)$ $\lambda_3^{'(2)} = 2^{3n}(2^n - 1)(2^n - 2)$ $.(2^{3n} - 7.2^{2n} + 18.2^n - 16)$
	$\lambda_2^{''(2)} = \lambda_2'' = 2^{3n}.(2^n - 1)$	various λ_3'' values
	$\downarrow$ $Adv_2 \leq \frac{2}{2^n - 1}$	$\downarrow$ $Adv_3 \leq \frac{2[3.2^{3n} - 18.2^{2n} + 36.2^n - 8]^{1/3}}{(2^n - 1)(2^n - 2)}$ $Adv_3 \lesssim \frac{2{,}88}{2^n}$

[15] Jacques Patarin. A Proof of Security in $O(2^n)$ for the Xor of Two Random Permutations . In Reihaneh Safavi-Naini, editor, *ICITS 2008*, volume 5155 of *Lecture Notes in Computer Science*, pages 232–248. Springer-Verlag, 2008. An extended version is also on eprint.

[16] F. Salzborn and G. Szekeres. A Problem in Combinatorial Group Theory. *Ars Combinatoria*, 7:3–5, 1979.

Appendices

A Examples: $\lambda_1,\ \lambda_2,\ \lambda_3,\ \lambda_2^{'(2)},\ \ \lambda_2^{'(2)}(\Psi)$

As examples, we present here the exact values for $\lambda_1,\ \lambda_2,\ \lambda_3$. We will see that they follow the values given in table 1. From λ_m we get a majoration for Adv_m by using the inequality (5.6).

A.1 Computation of λ_1

$$\lambda_1 =_{def} \text{Number of } (f_1, g_1, h_1) \text{ with } f_1, g_1, h_1 \in I_n$$

Therefore $\underline{\lambda_1 = 2^{3n}}$.

A.2 Computation of λ_2

Computation of λ_2 from (7.2)

$\lambda_2 =_{def}$ Number of $(f_1, g_1, h_1), (f_2, g_2, h_2)$ such that $f_2 \neq f_1,\ g_2 \neq g_1,\ h_2 \neq h_1,\ f_2 \oplus g_2 \oplus h_2 \neq f_1 \oplus g_1 \oplus h_1$

From the general formula (7.1) or (7.2) of Section 7, we have (with $\alpha = 1$):

$$\lambda_2 = [2^{3n} - 4 \cdot 2^{2n} + 6 \cdot 2^n - 3]\lambda_1 + 0$$

(here $[\lambda_1'] = 0$ since we have only one indice and in X we must have at least two indices).

$$\underline{\lambda_2 = [2^{3n} - 4 \cdot 2^{2n} + 6 \cdot 2^n - 3] \cdot 2^{3n}}$$

Computations of λ_2 from the β_i equations

$$\lambda_2 = 2^{3n}\lambda_1 - \sum_{i=1}^{4}|B_i| + \sum_{i<j}|B_i \cap B_j| - \sum_{i<j<k}|B_i \cap B_j \cap B_k| + \sum_{i<j<k<l}|B_i \cap B_j \cap B_k \cap B_l|$$

1 equation: $\sum_{i=1}^{4}|B_i| = 4 \cdot 2^{2n}\lambda_1$.
2 equations: $\sum_{i<j}|B_i \cap B_j| = 6 \cdot 2^n\lambda_1$.
3 equations: $\sum_{i<j<k}|B_i \cap B_j \cap B_k| = 4\lambda_1$.
4 equations: $\sum_{i<j<k<l}|B_i \cap B_j \cap B_k \cap B_l| = \lambda_1$.
Therefore $\lambda_2 = (2^{3n} - 4 \cdot 2^{2n} + 6 \cdot 2^n - 3)\lambda_1$ (as expected we obtain the same result as above).

A.3 Computation of λ_3 and $\lambda_2^{'(2)}$

Computation of λ_3 from (7.2)

From the general formulas (7.1) and (7.2), we have (with $\alpha = 2$):

$$\lambda_3 = (2^{3n} - 8 \cdot 2^{2n} + 24 \cdot 2^n - 30)\lambda_2 + 6\lambda_2^{'(2)}$$

(here $\lambda_2^{'(3)} = 0$ and $\lambda_2^{'(4)} = 0$ since we have here only 2 indices) where $\lambda_2^{'(2)}$ is the number of $(f_1, g_1, h_1),\ (f_2, g_2, h_2)$ such that $f_2 \neq f_1,\ g_2 \neq g_1,\ h_2 \neq h_1,\ f_2 \oplus g_2 \oplus h_2 \neq f_1 \oplus g_1 \oplus h_1$ and $f_1 \oplus g_1 = f_2 \oplus g_2$ (all the other equations X of the type $\lambda_2^{'(2)}$ give the same value $\lambda_2^{'(2)}$). When f_1, g_1, h_1 are fixed (we have 2^{3n} possibilities) then we will choose $f_2 \neq f_1$, $h_2 \neq h_1$, and $g_2 = f_1 \oplus f_2 \oplus g_1$ (so we have $g_2 \neq g_1$ and $f_2 \oplus g_2 \oplus h_2 \neq f_1 \oplus g_1 \oplus h_1$). Therefore $\underline{\lambda_2^{'(2)} = 2^{3n} \cdot (2^n - 1)^2}$ and the exact value of λ_3 is:

$$\lambda_3 = (2^{3n} - 8 \cdot 2^{2n} + 24 \cdot 2^n - 30)\lambda_2 + 6 \cdot 2^{3n} \cdot (2^n - 1)^2$$

(with $\lambda_2 = (2^{3n} - 4 \cdot 2^{2n} + 6 \cdot 2^n - 3) \cdot 2^{3n}$ as seen above).
This gives

$$\underline{\lambda_3 = 2^{9n} - 12 \cdot 2^{8n} + 62 \cdot 2^{7n} - 177 \cdot 2^{6n} + 294 \cdot 2^{5n} - 264 \cdot 2^{4n} + 96 \cdot 2^{3n}}$$

Computation of $\lambda_2^{'(2)}$ from the β_i equations The β_i equations have been defined in section 6. (We proceed here as in section 10 or as in Theorem 5 "General induction formula").

$$\lambda_2' = 2^{2n}\lambda_1 - \sum_{i=1}^{4}|B_i'| + \sum_{i<j}|B_i' \cap B_j'| - \sum_{i<j<k}|B_i' \cap B_j' \cap B_k'| + \sum_{i<j<k<l}|B_i' \cap B_j' \cap B_k' \cap B_l'|$$

Here X is: $f_1 \oplus f_2 = g_1 \oplus g_2$

• $X + 1$ **equation**.

$$\sum_{i=1}^{4}|B_i'| = 4 \cdot 2^n\lambda_1$$

• $X + 2$ **equations**. If the 2 equations β_i are ($f_1 = f_2$ and $g_1 = g_2$), or ($h_1 = h_2$ and $f_1 \oplus g_1 \oplus h_1 = f_2 \oplus g_2 \oplus h_2$), then X is the Xor of these equations. Therefore

$$\sum_{i<j}|B_i' \cap B_j'| = 4 \cdot \lambda_1 + 2 \cdot 2^n\lambda_1$$

• $X+3$ **equations**. X is always a consequence of the 3 equations, $\sum_{i<j<k}|B'_i\cap B'_j\cap B'_k| = 4\lambda_1$.
• $X+4$ **equations**. $\sum_{i<j<k<l}|B'_i\cap B'_j\cap B'_k\cap B'_l| = \lambda_1$.
Therefore

$$\lambda_2^{'(2)} = (2^{2n} - 4\cdot 2^n + 4 - 2\cdot 2^n - 4 + 1)\lambda_1$$

$$\lambda_2^{'(2)} = (2^{2n} - 2\cdot 2^n + 1)\lambda_1$$

(as expected we obtain the same result as above).

Remark. Here

$$\frac{2^n\lambda_2^{'(2)}}{\lambda_2} = \frac{1-\frac{2}{2^n}+\frac{1}{2^{2n}}}{1-\frac{4}{2^n}+\frac{6}{2^{2n}}-\frac{3}{2^{3n}}} = 1+\frac{2}{2^n}+\frac{3}{2^{2n}}+O(\frac{1}{2^{3n}})$$

and here $\epsilon_2 = \frac{2}{2^n}+\frac{3}{2^{2n}}+O(\frac{1}{2^{3n}})$. Therefore we see that in $\dfrac{2^n\lambda'_\alpha}{\lambda_\alpha}$, we have sometimes a term in $O(\frac{1}{2^n})$. However this is exceptional: here $f_1\oplus g_1 = f_2\oplus g_2$ is the Xor of the conditions $f_1\neq f_2$ and $g_1\neq g_2$, or of the conditions $h_1\neq h_2$ and $f_2\oplus g_2\oplus h_2 \neq f_1\oplus g_1\oplus h_1$. (or, this equation X is not strong, with the definition of "strong" given in section 7). Moreover here we have only 2 indices.

A.4 Computation of $\lambda_2^{'(2)}(\Psi)$

Let $\Psi\in I_n$, $\lambda_2^{'(2)}(\Psi)$ is by definition the number of $(f_1,g_1,h_1),(f_2,g_2,h_2)$ such that $f_2\neq f_1,\ g_2\neq g_1,\ h_1\neq h_2,\ f_2\oplus g_2\oplus h_2\neq f_1\oplus g_1\oplus h_1$, and this equation X is satisfied:
$X: f_1\oplus g_1 = f_2\oplus g_2\oplus\Psi$.
When $\Psi=0$, $\lambda_2^{'(2)}(\Psi)$ is simply denoted $\lambda_2^{'(2)}$ and this value is given above (in A.3). We will assume here that $\Psi\neq 0$.
First Computation
For f_1,g_1,h_1 we have 2^{3n} possibilities. Now from X, $f_1\neq f_2$ and $g_1\neq g_2$, we see that $f_2\notin\{f_1,f_1\oplus\Psi\}$ and $g_2\notin\{g_1,g_1\oplus\Psi\}$. Therefore, if $\Psi\neq 0$, we have: $\lambda_2^{'(2)}(\Psi) = 2^{3n}\cdot(2^n-2)^2$.
Second Computation
With the same notations as in (A.3) we have:

$$\lambda'_2(\psi) = 2^{2n}\lambda_1 - \sum_{i=1}^{4}|B'_i| + \sum_{i<j}|B'_i\cap B'_j| - \sum_{i<j<k}|B'_i\cap B'_j\cap B'_k|$$

$$+\sum_{i<j<k<l}|B'_i\cap B'_j\cap B'_k\cap B'_l|$$

• $X+1$ **equation:** $\sum_{i=1}^4|B'_i| = 4\cdot 2^n\lambda_1$ since 2 variables (among f_2,g_2,h_2) are fixed.
• $X+2$ **equations:** $\sum_{i<j}|B'_i\cap B'_j| = 4\cdot\lambda_1$ if $\Psi\neq 0$ since among the 6 possibilities, 4 fix the variables and 2 are impossible (they give $\Psi=0$).
• $X+3$ **equations and** $X+4$ **equations:** 0 solutions, since by Xoring we get $\Psi=0$.
Therefore: if $\Psi\neq 0$, we have: $\lambda_2^{'(2)}(\Psi) = (2^{2n}-4\cdot 2^n+4)\lambda_1$. As expected, we obtain the same value with the first and the second computation. Moreover, we can check that: $\sum_{\Psi\in I_n}\lambda_2^{'(2)}(\Psi) = \lambda_2$

B λ_α as a polynomial in 2^n

We have seen above that λ_1, λ_2 and λ_3 are polynomials in 2^n. We will see now that this is the case for any λ_α.
λ_α is by definition the number of $(f_1, g_1, h_1, \ldots, f_\alpha, g_\alpha, h_\alpha) \in I_n^{3\alpha}$ such that

$$\forall i, j,\ 1 \le i < j \le \alpha:\ f_i \neq f_j,\ g_i \neq g_j\ \ h_i \neq h_j,\ f_i \oplus g_i \oplus h_i \neq f_j \oplus g_j \oplus h_j$$

We have here $4 \cdot \frac{\alpha(\alpha-1)}{2} = 2\alpha^2 - 2\alpha$ conditions. Let $\beta_1, \beta_2, \ldots, \beta_{2\alpha^2-2\alpha}$ be these equalities (for example β_1 is $f_1 = f_2$).
$\forall i, 1 \le i \le 2\alpha^2 - 2\alpha$, let $B_i =$ the set of all $(f_1, g_1, h_1, \ldots, f_\alpha, g_\alpha, h_\alpha) \in I_n^{3\alpha}$ such that the equation β_i is satisfied. Then $\lambda_\alpha = 2^{3\alpha n} - |\cup_{i=1}^{2\alpha^2-2\alpha} B_i|$ (1).
For any sets we have:

$$|\cup_{i=1}^{k} B_i| = \sum_{i=1}^{k} |B_i| - \sum_{i<j} |B_i \cap B_j| + \sum_{i<j<k} |B_i \cap B_j \cap B_k| + \ldots + (-1)^{k+1} |B_i \cap B_2 \cap \ldots \cap B_k| \quad (2)$$

Moreover $|B_{i_1} \cap B_{i_2} \cap \ldots \cap B_{i_l}|$ is the number of $(f_1, g_1, h_1, \ldots, f_\alpha, g_\alpha, h_\alpha) \in I_n^{3\alpha}$ such that l linear equalities are satisfied. If these equalities are not compatible, then $|B_{i_1} \cap B_{i_2} \cap \ldots \cap B_{i_l}| = 0$. If these equalities are compatible, and if at most μ of them are independent, then $|B_{i_1} \cap B_{i_2} \cap \ldots \cap B_{i_l}| = 2^{(3\alpha-\mu)n}$ (3). (Since μ variables are fixed and the other are independent here). Therefore, from (1), (2) and (3) we see that λ_α is a polynomial in 2^n. We also see that this polynomial is of degree 3α, and that it has alternatively the sign $+$ and the sign $-$ when the monomials are ordered with decreasing degrees.

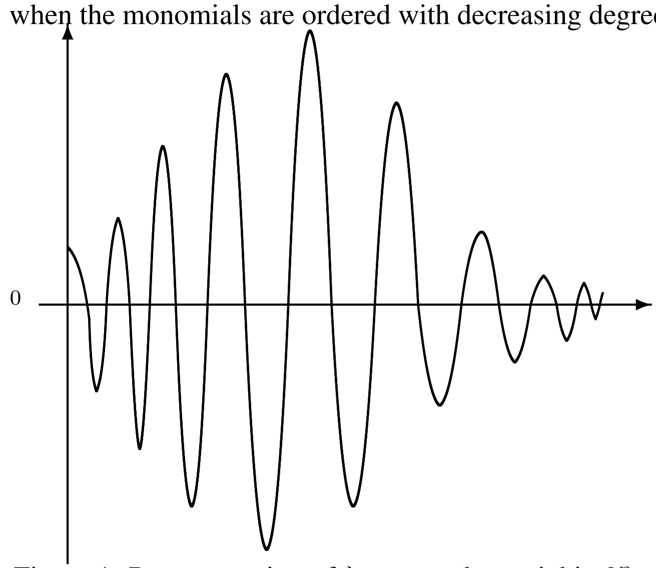


Figure 1: Representation of λ_α as a polynomial in 2^n.

C Summary of our notation and of our General Proof Strategy

In this Appendix C we will summarize the proof strategy and the main notations used in this paper.

- m and n are two integers. $I_n = \{0, 1\}^n$. (from a cryptographic point of view, m will be the number of queries, and n is the number of bits of the inputs and outputs of each query).

- H_m (cf section 3) denotes the number of $(f,g) \in B_n^2$ such that $\forall i,\ 1 \le i \le m,\ (f \oplus g)(a_i) = b_i$. H_m is a compact notation for $H_m(b_1, b_2, \dots, b_m)$.
- h_m (cf section 3) denotes the number of $(P_1, P_2, \dots, P_m, Q_1, Q_2, \dots, Q_m) \in I_n^{2m}$ such that: the P_i are pairwise distinct, the Q_i are pairwise distinct, and: $\forall i,\ 1 \le i \le m,\ P_i \oplus Q_i = b_i$. h_m is a compact notation for $h_m(b_1, b_2, \dots, b_m)$. ($H_m$ and h_m are equal up to a multiplicative constant: $H_m = h_m \cdot \frac{|B_n|^2}{(2^n(2^n-1)\dots(2^n-m+1))^2}$, cf formula $\sharp$ of section 3).
- λ_m (cf section 5) denotes the number of sequences $(f_i, g_i, h_i),\ 1 \le i \le m,\ f_i, g_i, h_i \in I_n$ such that: the f_i are pairwise distinct, the g_i are pairwise distinct, the h_i are pairwise distinct, and the $f_i \oplus g_i \oplus h_i$ are pairwise distinct, $1 \le i \le m$.
- U_m denotes $\frac{(2^n(2^n-1)\dots(2^n-m+1))^4}{2^{nm}}$

In Part I (sections 3,4,5), by the analysis of $E(H)$ and $\sigma(H)$ (i.e. "H_σ technique") we have proved that for all CPA-2 attacks ϕ with m queries:

$$Adv_\phi^{PRF} \le 2(\frac{\lambda_m}{U_m} - 1)^{1/3} \quad \text{cf (5.6)}$$

Therefore, the general proof strategy used in this paper was to study the λ_m values and to show that: when $m \ll 2^n,\ \lambda_m \simeq U_m\ (C1)$. (In **[?]**; a slightly different proof strategy called "standard H technique" will be used, with similar, but slightly different results).
In order to prove $(C1)$, we proceed in this paper with what we call the "usual proof strategy in Mirror Theory" or the "colored proof strategy". ("Mirror Theory" is the theory that analyses the number of solutions of sets of affine equalities (=) and affine non equalities ($\neq$) in finite fields). Essentially the two main ideas of this "colored proof strategy" are:

1. To compare $\frac{\lambda_{\alpha+1}}{\lambda_\alpha}$ with $\frac{U_{\alpha+1}}{U_\alpha}$ and to use

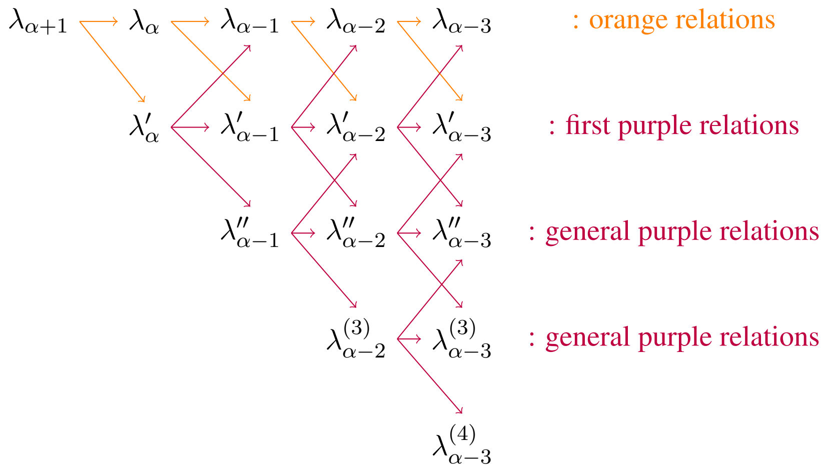
   $$\lambda_\alpha = \frac{\lambda_\alpha}{\lambda_{\alpha-1}} . \frac{\lambda_{\alpha-1}}{\lambda_{\alpha-2}} . \frac{\lambda_{\alpha-2}}{\lambda_{\alpha-3}} \dots \frac{\lambda_2}{\lambda_1}\lambda_1$$

   instead of studying λ_α globally.
2. To look carefully if the affine equations that will appear in the analysis of $\frac{\lambda_{\alpha+1}}{\lambda_\alpha}$ are independent, consequences, or in contradiction with the linear equalities in λ_α.

More precisely, here, with λ_α values, this "colored proof strategy" is this one:

1. We get an equation (called the "orange equation") that evaluates $\lambda_{\alpha+1}$ from λ_α and λ'_α (where $\lambda'_\alpha(X)$ denotes the number of solutions that satisfy the conditions λ_α plus one equality X: $f_i \oplus g_j \oplus h_k = f_l \oplus g_l \oplus h_l$, and where λ'_α denotes any value of $\lambda'_\alpha(X)$ when this equality X is linearly independent with the non equalities of λ_α). This was done in section 7 of this paper.
2. We get an equation (called the "first purple equation") that evaluates λ'_α from $\lambda_{\alpha-1}$, $\lambda'_{\alpha-1}$ and $\lambda''_{\alpha-1}$ (where in $\lambda''_{\alpha-1}$ we have introduced two extra and independent affine equations from the $\lambda_{\alpha-1}$ conditions). It is sometimes interesting (since it sometimes simplifies the analysis) to introduce a constant Ψ in the affine equations X.

Figure 2: General view of the "colored proof strategy" used in this paper



3. We get the equations (called "all purple equations") that evaluate $\lambda^{(d)}_\alpha$ from $\lambda^{(d-1)}_{\alpha-1}$, $\lambda^{(d)}_{\alpha-1}$, and $\lambda^{(d+1)}_{\alpha-1}$, (where in $\lambda^{(d)}_{\alpha-1}$, we have introduced d extra and independent affine equations from the $\lambda_{\alpha-1}$ equations).

4. Now, from these evaluations we are able to compare $\frac{\lambda_{\alpha+1}}{\lambda_\alpha}$ with $\frac{U_{\alpha+1}}{U_\alpha}$ and therefore λ_α from U_α. This can be done either with the constant Ψ (by looking for the possible deviation) or with $\Psi = 0$ (by evaluating λ_α).

D An induction formula on $\lambda^{'(4)}_\alpha$ ("First purple equation" on $\lambda^{'(4)}_\alpha$)

The values $\lambda^{'(4)}_\alpha$ have been introduced in section 7. By definition, $\lambda^{'(4)}_{\alpha+1}$ is the number of sequences (f_i, g_i, h_i), $1 \le i \le \alpha+1$, such that

1. The conditions $\lambda_{\alpha+1}$ are satisfied.

2. This equation X is satisfied:

$$\underline{X : f_{\alpha+1} \oplus g_{\alpha+1} \oplus h_{\alpha+1} = f_1 \oplus g_2 \oplus h_3}$$

(there we have chosen the indices $\alpha+1$, 1, 2, 3 but all other choices of 4 distinct indices give the same result $\lambda^{'(4)}_{\alpha+1}$ due to the symmetries of the conditions $\lambda_{\alpha+1}$. For example with $X : h_{\alpha+1} = f_1 \oplus g_1 \oplus h_1 \oplus f_2 \oplus g_3$, we would get exactly the same value $\lambda^{'(4)}_{\alpha+1}$).

In this section, we will compute $\lambda^{'(4)}_{\alpha+1}$ from λ_α and other values with indices less than or equal to α.

For each i, $1 \le i \le 4\alpha$, we will denote by B'_i the set of

$$(f_1, \ldots, f_{\alpha+1}, g_1, \ldots, g_{\alpha+1}, h_1, \ldots, h_{\alpha+1})$$

that satisfy the conditions λ_α and that satisfy the equation β_i, and the equation X (the β_i equations have been defined in Section 6). Therefore we have:

$$\lambda_{\alpha+1}^{'(4)} = 2^{2n}\lambda_\alpha - |\cup_{i=1}^{4\alpha} B_i'|$$

We will proceed here exactly as in section 6, but with the sets B_i' instead of the sets B_i. Since 5 equations β_i are always incompatible with the conditions λ_α, we have:

$$\lambda_{\alpha+1}^{'(4)} = 2^{2n}\lambda_\alpha - \sum_{i=1}^{4\alpha}|B_i'| + \sum_{i<j}|B_i' \cap B_j'| - \sum_{i<j<k}|B_i' \cap B_j' \cap B_k'| + \sum_{i<j<k<l}|B_i' \cap B_j' \cap B_k' \cap B_l'|$$

• $X+1$ **equation**.

Case 1: β_i is not an equation in $f_{\alpha+1} \oplus g_{\alpha+1} \oplus h_{\alpha+1}$. Then X and β_i will fix two variables among $f_{\alpha+1}, g_{\alpha+1}, h_{\alpha+1}$ from the other variables f_i, g_i, h_i. Therefore:

$$|B_i'| = 2^n\lambda_\alpha$$

Case 2: β_i is $f_{\alpha+1} \oplus g_{\alpha+1} \oplus h_{\alpha+1} = f_l \oplus g_l \oplus h_l$, for a value $l \leq \alpha$. Then $|B_i'| = 2^{2n}[\lambda_\alpha']$, where $[\lambda_\alpha']$ denotes the number of (f_i, g_i, h_i), $1 \leq i \leq \alpha$, that satisfy the conditions λ_α plus the equation Y: $f_l \oplus g_l \oplus h_l = f_1 \oplus g_2 \oplus h_3$. When $l \notin \{1,2,3\}$, we will denote $[\lambda_\alpha']$ by $\lambda_\alpha^{'(4)}$, and if $l \in \{1,2,3\}$, we will denote $[\lambda_\alpha'] = \lambda_\alpha^{'(3)}$. From Cases 1 and 2, we get:

$$-\sum_{i=1}^{4\alpha}|B_i'| = -3\alpha \cdot 2^n\lambda_\alpha - (\alpha-3)\cdot 2^{2n}\lambda_\alpha^{'(4)} - 3\cdot 2^{2n}\lambda_\alpha^{'(3)}$$

• $X+2$ **equations**.

Let β_i and β_j be these two equations.

Case 1: β_i and β_j are two equations in f, or in g, or in h, or in $f \oplus g \oplus h$. Then $|B_i' \cap B_j'| = 0$.

Remark. This value is not a problem since in the analog term for U_α, we get also 0 here.

Case 2: β_i and β_j are not in $f \oplus g \oplus h$ and we are not in Case 1. Then $|B_i' \cap B_j'| = \lambda_\alpha$ and here we have $3\alpha^2$ possibilities for the indices. (Remark: we can sometimes obtain here $f_{\alpha+1} = f_1$, or $g_{\alpha+1} = g_2$, or $h_{\alpha+1} = h_3$ by Xoring X, β_i and β_j).

Case 3: β_i is in $f \oplus g \oplus h$, but not β_j (or the opposite). (Here we have $3\alpha^2$ possibilities for the indices). For example β_i is

$$f_{\alpha+1} \oplus g_{\alpha+1} \oplus h_{\alpha+1} = f_l \oplus g_l \oplus h_l$$

for a value $l \leq \alpha$. Then $X \oplus \beta_i$ is: $f_l \oplus g_l \oplus h_l = f_1 \oplus g_2 \oplus h_3$. With the same notation as above for $X+1$ equations, $|B_i' \cap B_j'| = 2^n[\lambda_\alpha']$, where $[\lambda_\alpha'] = \lambda_\alpha^{'(4)}$ if $l \notin \{1,2,3\}$ and $[\lambda_\alpha'] = \lambda_\alpha^{'(3)}$ if $l \in \{1,2,3\}$. (Remark: if $l = 1$ for example, we get $g_1 \oplus h_1 = g_2 \oplus h_3$ and from β_j we cannot get here $g_1 = g_2$ or $h_1 = h_3$ since in β_j we have the index $\alpha+1$). Then from Cases 1, 2, 3, we get:

$$\sum_{i<j}|B_i' \cap B_j'| = 3\alpha^2\lambda_\alpha + (3\alpha^2 - 9\alpha)2^n\lambda_\alpha^{'(4)} + 9\alpha \cdot 2^n\lambda_\alpha^{'(3)}$$

.

• $X+3$ **equations**.

Let β_i, β_j and β_k be these three equations.

Case 1: If we have with β_i, β_j, β_k, two conditions in f, or two conditions in g, or two conditions in h, or two conditions in $f \oplus g \oplus h$, then $|B'_i \cap B'_j \cap B'_k| = 0$.

Case 2: X is a linear dependency of β_i, β_j, β_k. Then β_i, β_j, β_k are: $[f_{\alpha+1} = f_1, g_{\alpha+1} = g_2, h_{\alpha+1} = h_3]$ and we have here: $|B'_i \cap B'_j \cap B'_k| = \lambda_\alpha$. (Remark: here $[f_{\alpha+1} \oplus g_{\alpha+1} \oplus h_{\alpha+1} = f_1 \oplus g_1 \oplus h_1,\ g_1 = g_2,\ \text{and}\ h_1 = h_3]$ is not a solution since $g_1 = g_2$ and $h_1 = h_3$ are not equations in β_i, i.e. they do not have the index $\alpha + 1$).

Case 3: X with β_i, β_j, β_k create an impossibility (for example $f_i = f_j$ with $i \neq j$). Here we have $|B'_i \cap B'_j \cap B'_k| = 0$ and $3(\alpha - 1)$ possibilities for the indices. (Here it is easy to check that in β_i, β_j, β_k we have no equation in $f \oplus g \oplus h$ since in the equations β_i we always have the index $\alpha + 1$).

Case 4: In β_i, β_j, β_k, we have one equation in f, one equation in g and one equation in h (none in $f \oplus g \oplus h$) and we are not in Case 2 or Case 3 (we have here $\alpha^3 - 3\alpha + 2$ possibilities for the indices). Then $|B'_i \cap B'_j \cap B'_k| = [\lambda'_\alpha]$, and in most of the cases, we have $[\lambda'_\alpha] = \lambda_\alpha^{'(6)}$ (i.e. 6 different indices).

Remark. We will not need it for the main results , but we give more details here. Let us consider that $\beta_i, \beta_j, \beta_k$ are $f_{\alpha+1} = f_i$, $g_{\alpha+1} = g_j$, $h_{\alpha+1} = h_k$, so with X we get:

$$f_1 \oplus g_2 \oplus h_3 = f_i \oplus g_j \oplus h_k\ (*) \text{ with } 1 \leq i \leq \alpha,\ 1 \leq j \leq \alpha,\ 1 \leq k \leq \alpha$$

We have α^3 possibilities for i, j, k. If we look what kind of equation $(*)$ all these α^3 possibilities give, we can show that we will obtain:

- **With 6 indices**: $(\alpha - 3)(\alpha - 4)(\alpha - 5) = \alpha^3 - 12\alpha^2 + 47\alpha - 60$ equations denoted $\lambda_\alpha^{'[6]}$ of Type: $f_1 \oplus f_2 \oplus g_3 \oplus g_4 \oplus h_5 \oplus h_6 = 0$ (the Type $f_1 \oplus g_1 \oplus h_1 \oplus f_2 \oplus g_2 \oplus h_2 \oplus g_3 \oplus g_4 \oplus h_5 \oplus h_6 = 0$ gives the same $\lambda_\alpha^{'[6]}$).
- **With 5 indices**: $9(\alpha-3)(\alpha-4) = 9\alpha^2 - 63\alpha + 108$ equations noted $\lambda_\alpha^{'[5]}$ of Type: $f_1 \oplus f_2 \oplus g_1 \oplus g_3 \oplus h_4 \oplus h_5 = 0$.
- **With 4 indices**: we will have here 4 families of equations:
  - $(3\alpha^2 - 15\alpha + 18)$ equations $\lambda_\alpha^{'[4,a]}$ of Type: $f_1 \oplus f_2 \oplus g_3 \oplus g_4 = 0$ (we also obtain the same $\lambda_\alpha^{'[4,a]}$ value for the Type: $f_1 \oplus f_2 \oplus g_1 \oplus g_2 \oplus g_3 \oplus g_4 \oplus h_1 \oplus h_2 = 0$).
  - $(12\alpha - 36)$ equations $\lambda_\alpha^{'[4,b]}$ of Type: $f_1 \oplus f_2 \oplus g_1 \oplus g_3 \oplus h_2 \oplus h_4$.
  - $(3\alpha - 9)$ equations $\lambda_\alpha^{'[4,c]}$ of Type: $f_1 \oplus f_2 \oplus g_1 \oplus g_2 \oplus h_3 \oplus h_4 = 0$ (we also obtain the same value $\lambda_\alpha^{'[4,c]}$ for the Type: $f_1 \oplus f_2 \oplus h_1 \oplus h_2 \oplus h_3 \oplus h_4 = 0$ or for the Type: $f_1 \oplus f_2 \oplus f_3 \oplus f_4 \oplus g_1 \oplus g_2 \oplus g_3 \oplus g_4 \oplus h_3 \oplus h_4 = 0$).
  - $(4\alpha - 12)$ equations $\lambda_\alpha^{'[4,d]}$ of Type $f_1 \oplus g_1 \oplus h_1 \oplus f_2 \oplus g_3 \oplus h_4 = 0$. (This case is simply $\lambda_\alpha^{'[4,d]} = \lambda_\alpha^{'(4)}$ as before).
- **With 3 indices**: We will have here 2 families of equations:
  - $(9\alpha - 12)$ equations $\lambda_\alpha^{'[3,a]}$ of Type: $f_1 \oplus f_2 \oplus g_1 \oplus g_3 = 0$, or of Type $f_1 \oplus f_2 \oplus g_1 \oplus g_2 \oplus h_1 \oplus h_3 = 0$ (same value as we can see by using the fact that f and $f \oplus g \oplus h$ play the same properties). This case is simply $\lambda_\alpha^{'[3,a]} = \lambda_\alpha^{'(3)}$ as before.
  - 2 equations $\lambda_\alpha^{'[3,b]}$ of Type: $f_1 \oplus f_2 \oplus g_1 \oplus g_3 \oplus h_2 \oplus h_3 = 0$.
- **With 2 indices**: 3 equations $\lambda_\alpha^{'[2]}$ of Type: $f_1 \oplus f_2 = g_1 \oplus g_2$
- **Special cases**
  - $(3\alpha - 3)$ impossibility of Type: $f_1 = f_2$
  - 1 equation of Type: $0 = 0$

If we add all these terms, we obtain α^3 terms as expected.

Case 5: In β_i, β_j, β_k, we have one $f \oplus g \oplus h$ and we are not in Case 1. (We have here $3\alpha^3$ possibilities for the indices and we cannot be in Case 2 or Case 3). Then $|B'_i \cap B'_j \cap B'_k| = [\lambda'_\alpha]$, and in most of the cases, we have here $[\lambda'_\alpha] = \lambda_\alpha^{'(4)}$ (i.e. 4 different indices).

Remark. Similarly, we can give more details here. Let us consider all the equations

$$f_l \oplus g_l \oplus h_l = f_1 \oplus g_2 \oplus h_3$$

We also have the equations $f_{\alpha+1} = f_i$ and $g_{\alpha+1} = g_j$, but they just fix $f_{\alpha+1}$ and $g_{\alpha+1}$. We have $1 \leq i \leq \alpha$, $1 \leq j \leq \alpha$ and $1 \leq l \leq \alpha$. If we look all the $3\alpha^3$ possibilities for these equations (the coefficient 3 comes here from no $h_{\alpha+1} = h_k$, no $f_{\alpha+1} = f_i$, or no $g_{\alpha+1} = g_j$), we obtain:

- **With 4 indices**: $3(\alpha-3)\alpha^2 = 3\alpha^3 - 9\alpha^2$ equations $\lambda_\alpha^{'[4,d]} (= \lambda^{'(4)})$
- **With 3 indices**: $9\alpha^2$ equations $\lambda_\alpha^{'[3,a]} (= \lambda^{'(3)})$

Then from cases 1, 2, 3, 4, 5 we get:

$$-\sum_{i<j<k} |B'_i \cap B'_j \cap B'_k| = -\lambda_\alpha - (4\alpha^3 - 3\alpha + 2)[\lambda'_\alpha]$$

where most of the $[\lambda'_\alpha]$ are $\lambda_\alpha^{'(6)}$ or $\lambda_\alpha^{'(4)}$.

• $X + 4$ **equations**.

If $|B'_i \cap B'_j \cap B'_k \cap B'_l| \neq 0$, we need to have one equation $f_{\alpha+1} = f_i$, one $g_{\alpha+1} = g_j$, one $h_{\alpha+1} = h_k$ and one $f_{\alpha+1} \oplus g_{\alpha+1} \oplus h_{\alpha+1} = f_l \oplus g_l \oplus h_l$. Then, with X, we obtain:

$$Y \quad \text{and} \quad Z: \quad f_l \oplus g_l \oplus h_l = f_i \oplus g_j \oplus h_k = f_1 \oplus g_2 \oplus h_3$$

Case 1: Y and Z give only one equation.

Then $(i = 1, j = 2, k = 3)$, or $(i = l, j = l, k = l)$ and we have α possibilities for l. (Therefore we have 2α possibilities for the indices). Then

$$|B'_i \cap B'_j \cap B'_k \cap B'_l| = \lambda_\alpha^{'(4)}$$

Case 2: X, λ_α and the 4 equations β_i are not compatible.

These cases are $(i = l, j = l, k \neq l)$, or $(j = l, k = l, i \neq l)$, or $(i = l, k = l, j \neq l)$, or $(j = 2, k = 3, i \neq 1)$, or $(i = 1, k = 3, j \neq 2)$ or $(i = 1, j = 2, k \neq 3)$ or $(i = j = k \neq l)$. So we have here $7\alpha(\alpha-1)$ possibilities for the indices.

Case 3: We are not in Case 1 or in Case 2. Then $|B'_i \cap B'_j \cap B'_k \cap B'_l| = [\lambda''_\alpha]$, where $[\lambda''_\alpha]$ denotes the number of (f_i, g_i, h_i), $1 \leq i \leq \alpha$, that satisfy the conditions λ_α plus the equations Y and Z: $f_l \oplus g_l \oplus h_l = f_i \oplus g_j \oplus h_k = f_1 \oplus g_2 \oplus h_3$.

Then from Cases 1, 2, 3, we get:

$$\sum_{i<j<k<l} |B'_i \cap B'_j \cap B'_k \cap B'_l| = 2\alpha\lambda_\alpha^{'(4)} + (\alpha^4 - 2\alpha - 7\alpha(\alpha-1))\lambda''_\alpha$$

Therefore the induction formula for $\lambda_{\alpha+1}^{'(4)}$ gives ("First purple equation on $\lambda_\alpha^{'(4)}$) :

$$\underline{\lambda_{\alpha+1}^{'(4)} = (2^{2n} - 3\alpha \cdot 2^n + 3\alpha^2 - 1)\lambda_\alpha + (-\alpha \cdot 2^{2n} + 3\alpha^2 \cdot 2^n - 4\alpha^3 + 5\alpha - 3)\lambda'_\alpha}$$

$$\underline{+(\alpha^4 - 7\alpha^2 + 5\alpha)\lambda''_\alpha} \quad (C1)$$

In this formula:

• The only term in $O(\alpha^4)$ in λ''_α is $\lambda_\alpha^{''(7)}$, i.e. is for $i, j, k, l, 1, 2, 3$ pairwise distinct with equations: $f_l \oplus g_l \oplus h_l = f_i \oplus g_j \oplus h_k = f_1 \oplus g_2 \oplus h_3$.

• The terms in $O(\alpha \cdot 2^{2n})$ or $O(\alpha^2 \cdot 2^n)$ or $O(\alpha^3)$ in λ'_α are $\lambda_\alpha^{'(4)}$ or $\lambda_\alpha^{'(6)}$. (From $X + 3$ equations we have two kinds of dominant terms).

So $\lambda_\alpha^{''(7)}$, $\lambda_\alpha^{'(4)}$ and $\lambda_\alpha^{'(6)}$ are needed. (We want something like: $\lambda_\alpha^{'(6)} = \frac{\lambda_\alpha}{2^n}(1 + O(\frac{1}{2^n}) + O(\frac{\alpha}{2^{2n}}))$ and $\lambda_\alpha^{''(7)} = \frac{\lambda_\alpha^{'(4)}}{2^n}(1 + O(\frac{1}{2^n}) + O(\frac{\alpha}{2^{2n}}))$. Now by induction from these terms, more general terms will appears. This is why we will establish properties on more general equations than λ_α and $\lambda_\alpha^{'(4)}$.

E Array of the dominant coefficient $\lambda^d_{\alpha-1}$

We have seen that in order to evaluate precisely $\lambda_{\alpha+1}$ from λ_α we need to evaluate λ'_α from λ_α. More precisely, we have seen that only one term in λ'_α was dominant: the term that we denoted $\lambda'^{(4)}_\alpha$ with 4 indices (typical $X : f_1 \oplus g_2 \oplus h_3 = f_4 \oplus g_4 \oplus h_4$).
Similarly, when we want to evaluate precisely λ'_α, we have seen a formula ("first purple equation") that gives λ'_α from $\lambda_{\alpha-1}$, $\lambda'_{\alpha-1}$ and $\lambda''_{\alpha-1}$. In this formula 2 terms in $\lambda'_{\alpha-1}$ will be dominant (with X with 4 or 6 indices) and one term in $\lambda''_{\alpha-1}$ will be dominant (with XY with 7 indices). This process will continue, with more precise evaluation at each level. The process, and the dominant terms that appear are shown in the array below. The generalization of the "first purple equation" is the "general purple equation" that evaluate(for any integer d) $\lambda^{d+1}_{\alpha+1}$ from λ^d_α, λ^{d+1}_α and λ^{d+2}_α. (This shown for example with the arrow in Table 2 for $\lambda''_{\alpha-2}$).

Table 2: Array of dominant terms

$\lambda_{\alpha+1}$	λ_α	$\lambda_{\alpha-1}$	$\lambda_{\alpha-2}$	$\lambda_{\alpha-3}$	...
	λ'_α X: 4 indices	$\lambda'_{\alpha-1}$ X: 4 or 6 indices	$\lambda'_{\alpha-2}$ X: 4,6 or 8 indices	$\lambda'_{\alpha-3}$ X: 4,6,8 or 10 indices	... ...
		$\lambda''_{\alpha-1}$ XY: 7 indices	$\lambda''_{\alpha-2}$ XY: 7 or 9 indices	$\lambda''_{\alpha-3}$ XY: 7,9 or 11 indices	... ...
			$\lambda'''_{\alpha-2}$ XYZ: 10 indices	$\lambda'''_{\alpha-3}$ XYZ: 10 or 12 indices	... ...
				$\lambda^4_{\alpha-3}$ $XYZT$: 13 indices	... ...

In this figure we see that for the term $\lambda^d_{\alpha-i}$ we need at most $(3i+4)-(i+1-d)$ indices $= 2i+d+3$ indices, and that we need only values d such that $d \leq i+1$. Therefore, if we denote by χ the number of indices in the equation (i.e. in X or XY or XYZ etc) of these terms, we always have: $\chi \leq 3i+4$. We can also notice that all these dominant terms $\lambda^d_{\alpha-i}$ are strong.

F Proof of a "coefficients H" Theorem

We present here a proof in English of a Theorem published in French in 1991 in J.Patarin PhD Thesis p.27(see [12]). This result was used in various papers (in Europe and Japan for example) but no English version of the proof was published so far. The corollary in the case of the Xor of two random permutations is also presented here.

Theorem 6 *Let k be an integer. Let K be a set of k-uples of functions $(f_1,\dots,f_k)$. Let G be an application of $K \to F_n$ (Therefore G is a way to design a function of F_n from k-uples $(f_1,\dots,f_k)$ of K). Let α and β be real numbers, $\alpha \geq 0$ and $\beta \geq 0$. Let $\mathcal{E}$ be a subset of I_n^m such that $|\mathcal{E}| \geq (1-\beta)\cdot 2^{nm}$.*

If:

1) For all sequences a_i, $1 \leq i \leq m$, of pairwise distinct elements of I_n and for all sequences b_i, $1 \leq i \leq m$, of $\mathcal{E}$ we have:

$$|H| \geq \frac{|K|}{2^{nm}}(1-\alpha)$$

where H denotes the number of $(f_1,\dots,f_k) \in K$ such that

$$\forall i,\ 1 \leq i \leq m,\ G(f_1,\dots f_k)(a_i) = b_i \quad (1)$$

Then

2) For every CPA-2 with m chosen plaintexts we have: $p \leq \alpha + \beta$ where $p = Adv_\phi^{PRF}$ denotes the advantage to distinguish $G(f_1, \ldots, f_k)$ when $(f_1, \ldots, f_k) \in_R K$ from a function $f \in_R F_n$ (2).

Proof of Theorem 5

(We follow here a proof, in French, of this Theorem in J.Patarin, PhD Thesis, 1991, Page 27).

Let ϕ be a (deterministic) algorithm which is used to test a function f of F_n. (ϕ can test any function f from $I_n \to I_n$). ϕ can use f at most m times, that is to say that ϕ can ask for the values of some $f(C_i)$, $C_i \in I_n$, $1 \leq i \leq m$. (The value C_1 is chosen by ϕ, then ϕ receive $f(C_1)$, then ϕ can choose any $C_2 \neq C_1$, then ϕ receive $f(C_2)$ etc). (Here we have adaptive chosen plaintexts). (If $i \neq j$, C_i is always different from C_j). After a finite but unbounded amount of time, ϕ gives an output of "1" or "0". This output (1 or 0) is noted $\phi(f)$.

We will denote by P_1^*, the probability that ϕ gives the output 1 when f is chosen randomly in F_n. Therefore

$$P_1^* = \frac{\text{Number of functions } f \text{ such that } \phi(f) = 1}{|F_n|}$$

where $|F_n| = 2^{n \cdot 2^n}$.

We will denote by P_1, the probability that ϕ gives the output 1 when $(f_1, \ldots, f_k) \in_R K$ and $f = G(f_1, \ldots, f_k)$. Therefore

$$P_1 = \frac{\text{Number of } (f_1, \ldots, f_k) \in K \text{ such that } \phi(G(f_1, \ldots, f_k)) = 1}{|K|}$$

We will prove:

(**"Main Lemma"**): For all such algorithms ϕ,

$$|P_1 - P_1^*| \leq \alpha + \beta$$

Then Theorem 1 will be an immediate corollary of this "Main Lemma" since Adv_ϕ^{PRF} is the best $|P_1 - P_1^*|$ that we can get with such ϕ algorithms.

Proof of the "Main Lemma"

Evaluation of P_1^*

Let f be a fixed function, and let $C_1, \ldots, C_m$ be the successive values that the program ϕ will ask for the values of f (when ϕ tests the function f). We will note $\sigma_1 = f(C_1), \ldots, \sigma_m = f(C_m)$. $\phi(f)$ depends **only** of the outputs $\sigma_1, \ldots, \sigma_m$. That is to say that if f' is another function of F_n such that $\forall i$, $1 \leq i \leq m$, $f'(C_i) = \sigma_i$, then $\phi(f) = \phi(f')$. (Since for $i < m$, the choice of C_{i+1} depends only of $\sigma_1, \ldots, \sigma_i$. Also the algorithm ϕ cannot distinguish f from f', because ϕ will ask for f and f' exactly the same inputs, and will obtain exactly the same outputs). Conversely, let $\sigma_1, \ldots, \sigma_n$ be m elements of I_n. Let C_1 be the first value that ϕ choose to know $f(C_1)$, C_2 the value that ϕ choose when ϕ has obtained the answer σ_1 for $f(C_1), \ldots$, and C_m the m^{th} value that ϕ presents to f, when ϕ has obtained $\sigma_1, \ldots, \sigma_{m-1}$ for $f(C_1), \ldots, f(C_{m-1})$. Let $\phi(\sigma_1, \ldots, \sigma_m)$ be the output of ϕ (0 or 1). Then

$$P_1^* = \sum_{\substack{\sigma_1, \ldots, \sigma_n \\ \phi(\sigma_1, \ldots \sigma_m) = 1}} \frac{\text{Number of functions } f \text{ such that } \forall i,\ 1 \leq i \leq m,\ f(C_i) = \sigma_i}{2^{n \cdot 2^n}}$$

Since the C_i are all distinct the number of functions f such that $\forall i,\ 1 \le i \le m,\ \ f(C_i) = \sigma_i$ is exactly $|F_n|/2^{nm}$. Therefore

$$P_1^* = \frac{\text{Number of outputs } (\sigma_1, \ldots, \sigma_m) \text{ such that } \phi(\sigma_1, \ldots \sigma_m) = 1}{2^{nm}}$$

Let N be the number of outputs $\sigma_1, \ldots, \sigma_m$ such that $\phi(\sigma_1, \ldots \sigma_m) = 1$. Then $P_1^* = \frac{N}{2^{nm}}$.

Evaluation of P_1

With the same notation $\sigma_1, \ldots, \sigma_n$, and $C_1, \ldots C_m$:

$$P_1 = \sum_{\substack{\sigma_1,\ldots,\sigma_n \\ \phi(\sigma_1,\ldots\sigma_m)=1}} \frac{\text{Number of } (f_1, \ldots, f_k) \in K \text{ such that } \forall i,\ 1 \le i \le m,\ G(f_1, \ldots, f_k)(C_i) = \sigma_i}{|K|} \quad (3)$$

Now (by definition of β) we have at most $\beta \cdot 2^{nm}$ sequences $(\sigma_1, \ldots, \sigma_m)$ such that $(\sigma_1, \ldots, \sigma_m) \notin \mathcal{E}$. Therefore, we have at least $N - \beta \cdot 2^{nm}$ sequences $(\sigma_1, \ldots, \sigma_m)$ such that $\phi(\sigma_1, \ldots \sigma_m) = 1$ and $(\sigma_1, \ldots, \sigma_m) \in E$ (4). Therefore, from (1), (3) and (4), we have

$$P_1 \ge \frac{(N - \beta \cdot 2^{nm}) \cdot \frac{|K|}{2^{nm}}(1-\alpha)}{|K|}$$

Therefore

$$P_1 \ge \big(\frac{N}{2^{nm}} - \beta\big)(1-\alpha)$$

$$P_1 \ge (P_1^* - \beta)(1-\alpha)$$

Thus $P_1 \ge P_1^* - \alpha - \beta$ (5), as claimed.

We now have to prove the inequality in the other side. For this, let P_0^* be the probability that $\phi(f) = 0$ when $f \in_R F_n$. $P_0^* = 1 - P_1^*$. Similarly, let P_0 be the probability that $\phi(f) = 0$ when $(f_1, \ldots, f_k) \in_R K$ and $f = G(f_1, \ldots, f_k)$. $P_0 = 1 - P_1$. We will have $P_0 \ge P_0^* - \alpha - \beta$ (since the outputs 0 and 1 have symmetrical hypothesis. Or, alternatively since we can always consider an algorithm ϕ' such that $\phi'(f) = 0 \Leftrightarrow \phi(f) = 1$ and apply (5) to this algorithm ϕ').

Therefore, $1 - P_1 \ge 1 - P_1^* - \alpha - \beta$, i.e. $P_1^* \ge P_1 - \alpha - \beta$ (6). Finally, from (5) and (6), we have: $|P_1 - P_1^*| \le \alpha + \beta$, as claimed.

Example of Application: Xor of two permutations

With $k = 2$, $K = |B_n|^2$ and $G(f_1, \ldots, f_k) = f_1 \oplus f_2$ we obtain immediately:

Theorem 7 *Let α and β be real numbers, $\alpha \ge 0$ and $\beta \ge 0$. Let $\mathcal{E}$ be a subset of I_n^m such that $|\mathcal{E}| \ge (1-\beta) \cdot 2^{nm}$.*

If:

1) For all sequences a_i, $1 \le i \le m$, of pairwise distinct elements of I_n and for all sequences b_i, $1 \le i \le m$, of $\mathcal{E}$ we have:

$$|H| \ge \frac{|B_n|^2}{2^{nm}}(1-\alpha)$$

where H denotes the number of $(f,g) \in B_n^2$ such that

$$\forall i,\ 1 \leq i \leq m,\ f \oplus g(a_i) = b_i$$

Then

2) For every CPA-2 with m chosen plaintexts we have: $p \leq \alpha + \beta$ where $p = Adv_{\phi}^{PRF}$ denotes the advantage to distinguish $f \oplus g$ when $(f,g) \in_R B_n^2$ from a function $h \in_R F_n$.