A Proof of Security in $O(2^n)$ for the Xor of Two Random Permutations

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 $m << 2^{\frac{2n}{3}}$. 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 $m << 2^{\frac{kn}{k+1}}$. In this paper we will prove that for $k = 2$, we have in fact already security when $m << O(2^n)$. Therefore we will obtain a proof of a similar result claimed in [2] (security when $m << O(2^n/n^{2/3})$). Moreover our proof is very different from the proof strategy suggested in [2] (we do not use Azuma inequality and Chernoff bounds for example), 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

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 (and the way that gives so far our best security result) is simply to Xor k independant 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 independant 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 explicitely. In this paper we will use a completely different proof strategy, based on the "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. 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 [7], 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 [9], it was proved that security in $O(2^n)$ was obtained not only for $k \to +\infty$, but already for $k = 7$. 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 $k = 2$.

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:* $\mathrm{Adv}_\phi^{\mathrm{PRF}} \leq O(\frac{m}{2^n})$ *where* $\mathrm{Adv}_\phi^{\mathrm{PRF}}$ *denotes the probability 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 $m << 2^n$. 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]) 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

We will use this general Theorem:

Theorem 2 *Let α and β be real numbers, $\alpha > 0$ and $\beta > 0$. Let E be a subset of I_n^n such that $|E| \geq (1 - \beta) \cdot 2^{nm}$. 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 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 $p = \text{Adv}_\phi^{\text{PRF}}$ denotes the probability to distinguish $f \oplus g$ when $(f,g) \in_R B_n^2$ from a function $h \in_R F_n$.*

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 [8] page 27 and a similar Theorem was used in [9] 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

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: $H \geq \frac{|B_n|^2}{2^{nm}}(1-\alpha)$ for a small value α (more precisely with $\alpha << O(\frac{m}{2^n})$). 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 "$H\sigma$ 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 $\sigma(H) << E(H)$ when $m << 2^n$. From Bienayme-Tchebichev Theorem, we have

$$Pr(|H - E(H)| \leq \alpha E(H)) \geq 1 - \frac{\sigma^2(H)}{\alpha^2 E^2(H)}$$

So

$$Pr[H \geq E(H)(1-\alpha)] \geq 1 - \frac{\sigma^2(H)}{\alpha^2 E^2(H)}$$

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

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

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 N 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 $x_i = f(a_i)$. Let N be the number of sequences x_i, $1 \leq i \leq m$, $x_i \in I_n$, such that:

3

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 $H = N \cdot \frac{|B_n|^2}{(2^n(2^n-1)...(2^n-m+1))}$. (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 $g(a_i) = b_i \oplus x_i$).

Thus we have
$$\text{Adv}_\phi^{\text{PRF}} \leq 2(\frac{\sigma(H)}{E(H)})^{2/3} = 2(\frac{\sigma(N)}{E(N)})^{2/3} \quad (3.1)$$

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

$$E(N) = \frac{(2^n(2^n-1)\dots(2^n-m+1))^2}{2^{nm}} \quad \text{and that} \quad \sigma(N) << E(N) \quad \text{when} \quad m << 2^n.$$

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 $\sigma(N)$ 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 slighty the bounds by using a more precise evaluation than the Bienayme-Tchebichev Theorem: instead of

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

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

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

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

4   Computation of $E(N)$

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 $N = \sum_{x \in J_n^m} \delta_x$. So we have $E(N) = \sum_{x \in J_n^m} E(\delta_x)$. 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

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

as expected.

5 First results on $V(N)$

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

$$V(N) = \sum_{x \in J_n^m} V(\delta_x) + \sum_{\substack{x,x' \in J_n^m \\ x \neq x'}} [E(\delta_x \, \delta_{x'}) - E(\delta_x) \, E(\delta_{x'})] \quad (5.1)$$

We will now study the 3 terms in (5.1), i.e. the terms in $V(\delta_x)$, the terms in $E(\delta_x \, \delta_{x'})$ and the terms in $E(\delta_x) \, E(\delta_{x'})$.

Terms in $V(\delta_x)$

$$V(\delta_x) = E(\delta_x^2) - (E(\delta_x))^2 = E(\delta_x) - (E(\delta_x))^2$$

$$V(\delta_x) = \frac{2^n(2^n-1)\dots(2^n-m+1)}{2^{nm}} - \frac{(2^n(2^n-1)\dots(2^n-m+1))^2}{2^{2nm}}$$

$$\text{So} \quad \sum_{x \in J_n^m} V(\delta_x) = \frac{(2^n(2^n-1)\dots(2^n-m+1))^2}{2^{nm}} - \frac{(2^n(2^n-1)\dots(2^n-m+1))^3}{2^{2nm}}$$

This term is less than $E(N)$ and therefore is much less than $E^2(N)$. (5.2)

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

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

$$\text{So} \quad \sum_{\substack{x,x' \in J_n^m \\ x \neq x'}} E(\delta_x) E(\delta_{x'}) = \frac{[2^n(2^n-1)\dots(2^n-m+1)-1][2^n(2^n-1)\dots(2^n-m+1)]^3}{2^{2nm}}$$

$$\simeq \frac{(2^n(2^n-1)\dots(2^n-m+1))^4}{2^{2nm}} = E^2(N) \quad (5.3)$$

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

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

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

$$\sum_{\substack{x,x' \in J_n^m \\ x \neq x'}} Pr_{b \in I_n^m} \left(\left\{ \begin{array}{ll} \text{The } x_i \text{ are pairwise distinct,} & 1 \leq i \leq m \\ \text{The } x_i \oplus b_i \text{ are pairwise distinct,} & 1 \leq i \leq m \end{array} \right. \right)$$

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

1. The x_i are pairwise distinct, $1 \leq i \leq m$.

2. The x_i' are pairwise distinct, $1 \leq i \leq m$.

3. The $x_i \oplus b_i$ are pairwise distinct, $1 \leq i \leq m$.

4. The $x_i' \oplus b_i$ are pairwise distinct, $1 \leq i \leq m$.

We have $A_m = \frac{\lambda_m}{2^{nm}}$ (5.4). Therefore from (5.1), (5.2), (5.3), (5.4), we have obtained:

$$V(N) \leq E(N) + E^2(N) - \frac{\lambda_m}{2^{nm}} \quad (5.5)$$

We want to prove that $V(N) << E^2(N)$. Therefore, our aim is to prove that

$$\lambda_m \simeq 2^{nm} \cdot E^2(N) = \frac{(2^n(2^n-1)\ldots(2^n-m+1))^4}{2^{nm}} \quad (5.6)$$

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, $1 \leq i \leq m$.

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 (5.6), we see that a suffisant condition is finally to prove that

$$\lambda_m = \frac{(2^n(2^n-1)\ldots(2^n-m+1))^4}{2^{nm}}(1+O(\frac{m}{2^n})) \quad (5.7)$$

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 (5.7) on the λ_m values. (We can notice that in $E(N)$ and $\sigma(N)$ we evalute 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 (5.7) is not obvious: we will need a few pages. However, fortunately, the mathematics that we will use are simple).

6 First results in λ_α

The values λ_α have been introduced in Section 5. Our aim is to prove (5.7), (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, in the Appendices, we will use this induction formula (7.2) to get our property on λ_α.

Let $U_\alpha = \frac{[2^n(2^n-1)\ldots(2^n-\alpha+1)]^4}{2^{n\alpha}}$. We have $U_{\alpha+1} = \frac{(2^n-\alpha)^4}{2^n}U_\alpha$.

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

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:

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

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

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

(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

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

and therefore we will get property (5.4):

$$\lambda_\alpha = U_\alpha(1 + 0(\frac{\alpha}{2^n}))$$

as wanted. Notice that to get here $0(\frac{\alpha}{2^n})$ we have used $0(\frac{\alpha}{2^{2n}})$ 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. $h_{\alpha+1} \notin \{h_1, \dots, h_\alpha\}$

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

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

First evaluation

When f_i, g_i, h_i values are fixed, $1 \le i \le \alpha$, 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, $1 \le i \le \alpha$, 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:

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

Therefore :

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

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

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

7

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

i.e. we get security until $\alpha^2 << 2^n$, i.e. until $\alpha << \sqrt{2^n}$. However, we want security until $\alpha << 2^n$ and not only $\alpha << \sqrt{2^n}$, 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

A more precise evaluation

For each i, $1 \leq i \leq 4\alpha$, 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 $i \neq j$ and $1 \leq \alpha, j \leq \alpha$, 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 $\sum_{i<j}|B_i \cap B_j| = 6\alpha^2 \cdot 2^n\lambda_\alpha$.

• **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 equations**.

8

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: $f_i \oplus g_j \oplus h_k = f_l \oplus g_l \oplus h_l$.

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 $\lambda'^{(4)}_\alpha$ this value of $\lambda'_\alpha(X)$. (The (4) here is to remember that we have exactly 4 indexes i, j, k, l).

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 $\lambda'^{(3)}_\alpha$ this value of $\lambda'_\alpha(X)$.

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 $\lambda'^{(2)}_\alpha$ this value of $\lambda'_\alpha(X)$.

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:

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

Therefore

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

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

We will denote by $[\lambda'_\alpha]$ any value of $\lambda'_\alpha(X)$ such that X is compatible with the conditions λ_α and such that X is not always true (X is not $0 = 0$). Then, from (7.1) we write

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

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. 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 $[\lambda'_\alpha]$ or for the values $\lambda'^{(4)}_\alpha$. This is what we will do in the Appendices.

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

Let $[\epsilon_\alpha] = \frac{2^n[\lambda'_\alpha]}{\lambda_\alpha} - 1$. Therefore, $[\lambda'_\alpha] = \frac{\lambda_\alpha}{2^n}(1 + [\epsilon_\alpha])$. From the analysis of the previous sections, we know that if we can prove that $\|[\epsilon_\alpha]\|$ is small, then $\mathrm{Adv}^{\mathrm{PRF}}_\phi$ will be small. Let evaluate more

precisely the links between $||\epsilon_\alpha||$ and $\mathrm{Adv}_\phi^{\mathrm{PRF}}$ that we have. From formula (7.2), we have:

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

with

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

Therefore, by using U_α of section 6 we have:

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

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

Therefore, with (8.1) we have

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

with

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

and

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

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

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

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

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

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

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

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

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

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

10

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, in the Appendices, we will present good bounds for $|[\epsilon_\alpha]|$. We will proceed progressively: first, in Appendix B, we will get 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, in Appendix D 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, in Appendix E, we will iterate the process in order to obtain security in $m << O(2^n)$ as wanted.

9   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 $\frac{m}{2^n} + O(n)(\frac{m}{2^n})^{3/2}$ 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'.

10   A simple property about the Xor of two permutations

We 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.

11   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, 7, 9], some results "above the birthday bound" on completely different cryptographic designs are also given). 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 combining the "coefficient H technique" of [8, 9] 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). 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.

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] 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.

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

[9] 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.

Appendices

A Examples: λ_1, λ_2, λ_3

As examples, we present here the exact values for λ_1, λ_2, λ_3.

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

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).

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

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).
Computation of λ_3 from (7.2)
From the general formulas (7.1) and (7.2), we have (with $\alpha = 2$):

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

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 $\lambda_2'^{(2)} = 2^{3n} \cdot (2^n - 1)^2$ 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).
Computation of $\lambda_\alpha'^{(2)}$ from the β_i equations

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

Here X is: $f_1 \oplus f_2 = g_1 \oplus g_2$
• $X + 1$ **equations**.

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

13

- $X+2$ **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

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

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

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

Remark. Here

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

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 $f_2 \oplus g_2 \oplus h_2 \neq f_1 \oplus g_1 \oplus h_1$. Moreover here we have only 2 indices.

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

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:*

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

14

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:

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

and

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

Therefore

$$2^{6n}\lambda_{\alpha-2}\left(1 - \frac{4(\alpha-1)}{2^n}\right)^2 \le \lambda_\alpha \le 2^{6n}\lambda_{\alpha-2} \quad (B1)$$

We will now evaluate $[\lambda'_\alpha]$ from $\lambda_{\alpha-2}$.

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

1. $f_{\alpha-1} \notin \{f_1, \ldots, f_{\alpha-2}\}$

2. $g_{\alpha-1} \notin \{g_1, \ldots, g_{\alpha-2}\}$

3. $h_{\alpha-1} \notin \{h_1, \ldots, h_{\alpha-2}\}$

4. $f_{\alpha-1} \oplus g_{\alpha-1} \oplus h_{\alpha-1} \notin \{f_1 \oplus g_1 \oplus h_1, \ldots, f_{\alpha-2} \oplus g_{\alpha-2} \oplus h_{\alpha-2}\}$

5. $f_\alpha \notin \{f_1, \ldots, f_{\alpha-1}\}$

6. $g_\alpha \notin \{g_1, \ldots, g_{\alpha-1}\}$

7. $h_\alpha \notin \{h_1, \ldots, h_{\alpha-1}\}$

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

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 $g_{\alpha-1}$:

$$h_{\alpha-1} \notin \{h_1, \ldots, h_{\alpha-2},$$

$$f_1 \oplus g_1 \oplus h_1 \oplus f_{\alpha-1} \oplus g_{\alpha-1}, \ldots, f_{\alpha-2} \oplus g_{\alpha-2} \oplus h_{\alpha-2} \oplus f_{\alpha-1} \oplus g_{\alpha-1},$$

$$f_1 \oplus g_1 \oplus h_1 \oplus f_{\alpha-3} \oplus g_{\alpha-2}, \ldots, f_{\alpha-2} \oplus h_{\alpha-2} \oplus f_{\alpha-3}\}$$

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

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

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

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

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

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

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

and

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

So

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

So we have:

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

and with $(B1)$ this gives:

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

So

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

as claimed.

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

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 $\left(\frac{8m^6}{2^{5n}(1 - \frac{12m}{2^n})}\right)^{1/3}$ in $O(\frac{m^2}{2^{5n/3}})$.

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

C An induction formula 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), $1 \leq i \leq \alpha + 1$, such that

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

2. This condition X is satisfied:

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

(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 $\lambda_{\alpha+1}$).

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

For each i, $1 \leq i \leq 4\alpha$, 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 conditions β_i, and the condition X (the β_i equations have been defined in Section 6). Therefore we have:

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

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:

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

- $X + 1$ **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 $\lambda'^{(4)}_\alpha$, and if $l \in \{1, 2, 3\}$, we will denote $[\lambda'_\alpha] = \lambda'^{(3)}_\alpha$. From Cases 1 and 2, we get:

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

- $X + 2$ **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 $|B'_i \cap B'_j| = 0$.

Case 2: β_i and β_j are not in $f \oplus g \oplus h$ and we are not in Case 1. Then $|B'_i \cap B'_j| = \lambda_\alpha$. (Here we have $3\alpha^2$ possibilities for the indices).

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

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

17

for a value $l \leq \alpha$. Then $X \oplus \beta_i$ is: $f_l \oplus g_l \oplus h_l = f_1 \oplus g_2 \oplus h_3$. Then, 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 $l \in \{1, 2, 3\}$. Then from Cases 1, 2, 3, we get:

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

.

- **$X + 3$ 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 dependancy 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: $|B_i' \cap B_j' \cap B_k'| = \lambda_\alpha$

Case 3: X is not a linear dependancy of β_i, β_j, β_k and in β_i, β_j, β_k, we have one equation in f, one equation in g and one equation in h (none in $f \oplus g \oplus h$). 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).

Case 4: X is not a linear dependancy of β_i, β_j, β_k and in β_i, β_j, β_k, we have one $f \oplus g \oplus h$ and we are not in Case 1. 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).

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

$$- \sum_{i < j < k} |B_i' \cap B_j' \cap B_k'| = -\lambda_\alpha - (4\alpha^3 - 1 - (3\alpha - 3))[\lambda_\alpha']$$

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 possibilitites for l. 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:

$$\sum_{i < j < k < l} |B_i' \cap B_j' \cap B_k' \cap B_l'| = 2\alpha \lambda_\alpha'^{(4)} + (\alpha^4 - 2\alpha - 7\alpha(\alpha - 1))[\lambda_\alpha'']$$

18

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

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

$$+(\alpha^4 - 7\alpha^2 + 5\alpha)[\lambda_\alpha''] \quad (C1)$$

In this formula:

• The only term in $O(\alpha^4)$ in $[\lambda_\alpha'']$ 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 $[\lambda_\alpha']$ are $\lambda_\alpha^{'(4)}$ or $\lambda_\alpha^{'(6)}$.

So $\lambda_\alpha^{''(7)}$ 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 $\lambda_\alpha^{'(4)}$ in the next Appendices.

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

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 B (Theorem 3) that $|[\epsilon_\alpha]| \leq \frac{8\alpha}{2^n \cdot (1 - \frac{8\alpha}{2^n})}$. 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:

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

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

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

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

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

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

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 proven (cf Appendix B):

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

we have

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

(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:

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

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:

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

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 $m << 2^{\frac{9n}{10}}$.

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

In Section 6, we have obtained security when $m << \sqrt{2^n}$. In Appendix B, we have obtained security when $m << 2^{\frac{5n}{6}}$. In Appendix D, we have obtained security when $m << 2^{\frac{9n}{10}}$. 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:

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

   for any integer k. ($[\mu]$ means that we have μ more equations, compatible and independant, 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 $m << 2^{\frac{kn}{k+1}}$, we see that the coefficient involved increases at most in 2^k. Therefore from (8.6) we will get:

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

(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,

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

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

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

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

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 [8]). 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 5 *Let k be an integer. Let K be a set of k-uples of functions $(f_1, \ldots, f_k)$. 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 $(f_1, \ldots, f_k)$ of K). Let α and β be real numbers, $\alpha > 0$ and $\beta > 0$. Let E be a subset of I_n^m such that $|E| \geq (1 - \beta) \cdot 2^{nm}$.*

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 E we have:

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

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

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

Then

2) For every CPA-2 with m chosen plaintexts we have: $p \leq \alpha + \beta$ where $p = Adv_\phi^{PRF}$ denotes the probability 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 $\forall i,\ 1 \le i \le m,\ f'(C_i) = \sigma_i$, 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 mth 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

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

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 $(\sigma_1, \ldots, \sigma_m) \notin E$. 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

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

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 symetrical 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.

23

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

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 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 $p = Adv_\phi^{PRF}$ denotes the probability to distinguish $f \oplus g$ when $(f, g) \in_R B_n^2$ from a function $h \in_R F_n$.