Le schéma d'Even-Mansour paramétrable : preuves de sécurité à l'aide de la technique des coefficients H

Benoît-Michel Cogliati

NNT : 2016SACLV064

Thèse de Doctorat
de
l'Université Paris-Saclay
préparée à
l'Université de Versailles Saint-Quentin en Yvelines

École doctorale n°580
Sciences et technologies de l'information et de la communication

Spécialité de doctorat : Informatique

Par

M. Benoît-Michel Cogliati

Le schéma d'Even-Mansour paramétrable : preuves de sécurité à l'aide de la technique des coefficients H

Thèse présentée et soutenue à Versailles, le 30 septembre 2016

Composition du Jury :

Louis Goubin, Professeur, Université de Versailles	Président du Jury
Pierre-Alain Fouque, Professeur, Université de Rennes I	Rapporteur
David Naccache, Professeur, ENS Paris	Rapporteur
Jacques Patarin, Professeur, Université de Versailles	Directeur
Jean-Sébastien Coron, Professeur assistant, Université de Luxembourg	Examinateur
Aline Gouget, Ingénieur de recherches, Gemalto	Examinatrice
Valérie Nachef, Maître de Conférence, Université de Cergy-Pontoise	Examinatrice
Yannick Seurin, Spécialiste en cryptographie, ANSSI	Invité

Remerciements

En premier lieu, je tiens à remercier Jacques Patarin, mon directeur de thèse, à qui je dois d'être tombé dans les preuves de sécurité en cryptographie symétrique. Durant cette thèse, j'ai pu admirer sa créativité, son intuition ainsi que sa ténacité, notamment face aux épineux problèmes de combinatoire rencontrés en cryptographie symétrique, que trop peu de chercheurs osent aborder. Je lui suis reconnaissant pour ses précieux conseils ainsi que pour l'autonomie et la liberté qu'il m'a accordées et qui m'ont permis de m'épanouir au cours de ces trois années.

Je tiens également à exprimer ma gratitude envers Yannick Seurin, avec lequel j'ai eu le plaisir d'avoir une fructueuse et enrichissante collaboration, pour le temps et l'énergie qu'il a consacrés à me guider, ainsi que pour sa disponibilité et son expertise. Sa capacité à identifier les sujets prometteurs, comme le schéma d'Even-Mansour paramétrable et les MACs de Wegman-Carter, a indubitablement contribué au succès de cette thèse.

Au cours de ces trois années, j'ai eu l'occasion de faire connaissance et d'échanger avec de nombreux collègues que je souhaite remercier : Sébastien Besnier, Ilaria Chilotti, Ninon Eyrolles, Cyril Hugounenq, Rodolphe Lampe, Francisco Vial, Christina Boura, Luca De Feo, Nicolas Gama, Michael Quisquater, Gaëtan Leurent, Mehdi Tibouchi,...

Je tiens à adresser des remerciements particuliers à Louis Goubin dont l'aide s'est avérée inestimable lors des épreuves administratives que nous avons pu rencontrer.

Je remercie chaleureusement Pierre-Alain Fouque et David Naccache qui ont accepté d'être les rapporteurs de cette thèse, ainsi que les autres relecteurs qui m'ont permis d'améliorer ce manuscrit, en particulier Jacques et Yannick. Je remercie également l'ensemble des membres du Jury d'être présent à cette soutenance. Je suis reconnaissant envers l'Université de Versailles qui m'a accueilli et a financé ma thèse.

Je tiens à remercier mes proches, dont le soutien sans faille m'a permis de mener à bien ce projet. Je remercie tout particulièrement mes parents, qui ont su cultiver ma curiosité intellectuelle et m'ont toujours encouragé dans mes choix. Pour terminer, je dédie cette thèse à Hélène, sans qui rien de tout ceci n'aurait été possible.

Table des matières

Table des figures

Notations

$\{0,1\}^n$	ensemble des chaînes de bits de taille n
$a \oplus b$	ou exclusif bit à bit des chaînes de bits de même taille a et b
$a \| b$	concaténation des deux chaînes de bits a et b
$(\mathcal{M})_q$	ensemble des $(x_1, \ldots, x_q) \in \mathcal{M}^q$ deux à deux distincts
$(a)_b$	pour $1 \leq b \leq a$, $(a)_b = a(a-1)\cdots(a-b+1)$ avec $(a)_0 = 1$
$\mathsf{Perm}(\mathcal{M})$	ensemble des permutations d'un ensemble $\mathcal{M}$
$\mathsf{Perm}(n)$	ensemble des permutations de $\{0,1\}^n$
$\mathsf{Func}(n)$	ensemble des fonctions de $\{0,1\}^n$ vers $\{0,1\}^n$
$\mathsf{BC}(\mathcal{K}, \mathcal{M})$	ensemble des algorithmes de chiffrement par blocs d'espace de messages $\mathcal{M}$ et d'espace de clés $\mathcal{K}$
$\mathsf{TP}(\mathcal{T}, n)$	ensemble des permutations paramétrables de $\{0,1\}^n$ d'espace de *tweaks* $\mathcal{T}$
$\mathsf{TBC}(\mathcal{K}, \mathcal{T}, \mathcal{M})$	ensemble des algorithmes de chiffrement par blocs paramétrables d'espace de messages $\mathcal{M}$, d'espace de *tweaks* $\mathcal{T}$ et d'espace de clés $\mathcal{K}$
$x \leftarrow_\$ \mu$	tirage d'un élément x selon la loi de probabilité μ
$x \leftarrow_\$ X$	tirage d'un élément x selon la loi uniforme sur X
$\Pr[x \leftarrow_\$ X : E]$	probabilité d'un évènement E lorsque $x \leftarrow_\$ X$
$\mathbb{E}[X]$	espérance d'une variable aléatoire X
$\|\mu - \nu\|$	distance statistique entre deux lois de probabilité μ et ν

Liste des abréviations

AES	*Advanced Encryption Standard*
CAESAR	Competition for Authenticated Encryption : Security, Applicability, and Robustness
cca	attaque adaptative à clair et chiffré choisi (*adaptive chosen ciphertext attack* en anglais)
cpa	attaque adaptative à clair choisi (*adaptive chosen plaintext attack* en anglais)
ε-AXU	ε-XOR universelle (ε-*Almost XOR Universal* en anglais)
EDM	construction de Davies-Meyer chiffrée (*Encrypted Davies-Meyer* en anglais)
EM	construction d'Even-Mansour
EWCDM	construction de Wegman-Carter chiffrée avec Davies-Meyer (*Encrypted Wegman-Carter with Davies-Meyer* en anglais)
MAC	code d'authentification de message (*Message Authentication Code* en anglais)
ncca	attaque non-adaptative à clair et chiffré choisi (*non-adaptive chosen ciphertext attack* en anglais)
ncpa	attaque non-adaptative à clair choisi (*non-adaptive chosen plaintext attack* en anglais)
OAEP	*Optimal Asymmetric Encryption Padding*
PRF	fonction pseudo-aléatoire (*PseudoRandom Function* en anglais)
PRP	permutation pseudo-aléatoire (*PseudoRandom Permutation* en anglais)
TEM	construction d'Even-Mansour paramétrable (*Tweakable Even-Mansour* en anglais)
XOR	ou exclusif bit à bit

Introduction

La cryptographie est une discipline dont l'existence remonte à l'Antiquité. D'après David Kahn [Kah96], les premières traces d'un usage limité de la cryptographie dans le but de protéger des informations ont été découvertes en Mésopotamie et remontent à plus de 3500 ans. Pendant de nombreux siècles, cette discipline a été presque exclusivement utilisée par les gouvernements, l'armée et les diplomates, afin de protéger les secrets d'état. Toutefois, au cours du XX^{e} siècle, l'apparition de l'informatique et l'essor fulgurant des télécommunications ont créé de nouveaux besoins en matière de cryptographie, notamment pour les entreprises et le grand public, afin de sécuriser les échanges de données. Ce développement a conduit à une profonde révolution de notre domaine. En effet, par le passé, les algorithmes utilisés étaient développés par tâtonnements, se voyant successivement attaqués, cassés puis réparés, sans plus de garantie de sécurité que l'absence temporaire d'attaque connue. Cette approche a petit à petit été jugée insuffisante et une nouvelle tendance a progressivement émergé : les cryptologues ont essayé de justifier la sécurité de leurs algorithmes, en fondant leurs arguments sur des raisonnements rigoureux. Les premières bases mathématiques sur lesquelles est fondée la cryptographie actuelle ont été posées en 1949 par Shannon dans l'article intitulé *Communication Theory of Secrecy Systems*. Il y prouve notamment que le chiffrement par masque jetable - aussi appelé chiffrement de Vernam - donne des garanties de sécurité très fortes : un attaquant ne peut tirer du chiffré d'un texte aucune autre information sur le texte clair que l'ordre de grandeur de sa longueur, et ce quelle que soit la puissance de calcul dont il dispose et sans aucune hypothèse. On parle ainsi de sécurité inconditionnelle. Les algorithmes inconditionnellement sûrs ne sont cependant pas parfaits : il sont extrêmement inefficaces et, de plus, certains types d'attaques, comme les attaques par canaux cachés, échappent à cette notion de sécurité. Le chiffrement de Vernam présente également un problème de malléabilité qui doit être géré par d'autres méthodes.

Afin d'obtenir des algorithmes plus efficaces, les cryptologues ont dû définir de nouvelles notions de sécurité qui, bien qu'elles soient plus faibles que la sécurité inconditionnelle, restent suffisamment fortes pour être dignes de confiance. Deux nouveaux types de preuves de sécurité sont apparus pour répondre à ces nouvelles problématiques. Le premier type est constitué des preuves dans le modèle standard. Dans ce modèle, les preuves de sécurité sont relatives : la sécurité d'un cryptosystème est réduite à un problème mathématique supposé difficile à résoudre. Cette difficulté correspond généralement au fait qu'il n'existe aucun algorithme capable de résoudre le

problème sous-jacent en un temps polynomial : il peut s'agir par exemple du problème de la factorisation, du logarithme discret, ou encore de mettre en défaut la nature pseudo-aléatoire d'un algorithme de chiffrement par blocs, comme l'algorithme AES. Dans de nombreux cas, le modèle standard s'avère délicat à manipuler. Il est alors courant, dans un second type de preuves, de substituer une primitive idéale à une partie du cryptosystème dont on cherche à prouver la sécurité : on parle ainsi de preuve dans un modèle idéalisé. Cette seconde catégorie de raisonnements est souvent utilisée pour prouver la sécurité de protocoles reposant sur des fonctions de hachage, comme par exemple le chiffrement OAEP [BR94], en remplaçant la fonction de hachage par une fonction choisie uniformément aléatoirement et accessible à l'attaquant. Ce modèle, dit de l'oracle aléatoire, a été utilisé pour la première fois en 1986 par Fiat et Shamir [FS86] puis formalisé en 1993 par Bellare et Rogaway [BR93]. Dans ce manuscrit, nous nous servirons d'un modèle idéalisé différent, dit de la permutation aléatoire, dans lequel on substitue des permutations publiques choisies uniformément aléatoirement à certaines composantes du cryptosystème. Les ressources remplacées étant très éloignées de primitives idéales, il va de soi que les preuves de sécurité dans un modèle idéalisé sont plus faibles que les preuves dans le modèle standard. Toutefois, si elles ne prouvent pas directement la sécurité du cryptosystème initial, elles permettent d'affirmer que sa structure est saine et que toute attaque à son encontre doit exploiter des caractéristiques des composantes que l'on a remplacées.

Les résultats présentés dans ce manuscrit appartiennent à ces deux dernières catégories et ont été démontrés grâce à une méthode de preuve particulière : la technique des coefficients H [Pat08b]. Cette technique a initialement été introduite par Jacques Patarin pour étudier la sécurité des schémas de Feistel équilibrés [Pat90, Pat91, Pat98, Pat03, Pat04] et déséquilibrés [Pat10], mais également pour prouver que le XOR de deux permutations pseudo-aléatoires donne une fonction pseudo-aléatoire sûre [Pat08a, Pat13]. Cette approche a ensuite été étendue par Chen et Steinberger afin d'analyser, dans le modèle de la permutation aléatoire, la sécurité du schéma d'Even-Mansour à plusieurs tours [CS14]. Quelques mois plus tard, Chen, Lampe, Lee, Steinberger et Seurin ont étudié une variante minimalisée de ce schéma, pour deux tours. Au cours de cette thèse, Rodolphe Lampe, Yannick Seurin et moi avons étendu cette analyse à une généralisation de cette construction, le schéma d'Even-Mansour paramétrable. Ces travaux ont donné lieu à trois publications [CS15b, CLS15, CS15a] qui feront l'objet de ce manuscrit. Dans un premier chapitre, nous commencerons par introduire quelques généralités sur les preuves de sécurité en cryptographie symétrique. Nous présenterons ensuite en détail le fonctionnement de la technique des coefficients H et l'illustrerons par deux cas concrets : l'étude, dans le modèle standard, de l'amplification de sécurité associée à la composition d'algorithmes de chiffrement par blocs équipés de clés indépendantes que Jacques Patarin, Yannick Seurin et moi avons effectuée [CPS14], suivie de celle de la construction EDM, une nouvelle méthode de conversion d'algorithmes de chiffrement par blocs en fonction pseudo-aléatoire introduite par Yannick Seurin et moi [CS16b]. Dans le chapitre deux, nous introduirons le schéma d'Even-Mansour ainsi que la notion d'algorithme de chiffrement par blocs paramétrable, avant de définir la construction d'Even-Mansour

paramétrable telle qu'elle a été introduite par Rodolphe Lampe, Yannick Seurin et moi [CLS15]. Les chapitres suivants seront consacrés à l'analyse de cette construction pour des fonctions de dérivation de clé fortement non linéaires, dans le chapitre trois, puis linéaires dans le chapitre quatre.

Au cours de cette thèse, Jacques Patarin, Rodolphe Lampe et moi avons également partiellement étendu l'analyse par Jacques Patarin du XOR de deux permutations aléatoires au cas d'un nombre plus élevé de permutations [CLP14], ce qui a permis d'améliorer significativement la borne de sécurité de [Luc00] tout en gardant une preuve très simple. Yannick Seurin et moi avons aussi étudié la résistance du schéma d'Even-Mansour aux attaques à clés choisies [CS15b] et aux attaques à clés connues [CS16b]. Enfin, Yannick Seurin et moi proposons et analysons une nouvelle variante des MACs de Wegman-Carter dérivée de la construction EDM présentée au premier chapitre que nous avons baptisée EWCDM (Encrypted Wegman-Carter with Davies-Meyer) [CS16a].

Liste de mes publications

Les articles exposés dans ce manuscrit apparaissent en gras.

[CS16a] **Benoît Cogliati, Yannick Seurin**
EWCDM : An Efficient, Beyond-Birthday Secure, Nonce-Misuse Resistant MAC
In *Advances in Cryptology - CRYPTO 2016*

[CS16b] Benoît Cogliati, Yannick Seurin
Strengthening the Known-Key Security Notion for Block Ciphers
In *Fast Software Encryption - FSE 2016*

[CS15a] **Benoît Cogliati, Yannick Seurin**
Beyond-Birthday-Bound Security for Tweakable Even-Mansour Ciphers with Linear Tweak and Key Mixing
In *Advances in Cryptology - ASIACRYPT 2015*

[CLS15] **Benoît Cogliati, Rodolphe Lampe, Yannick Seurin**
Tweaking Even-Mansour Ciphers
In *Advances in Cryptology - CRYPTO 2015*

[CS15b] **Benoît Cogliati, Yannick Seurin**
On the Provable Security of the Iterated Even-Mansour Cipher Against Related-Key and Chosen-Key Attacks
In *Advances in Cryptology - EUROCRYPT 2015*

[CPS14] **Benoît Cogliati, Jacques Patarin, Yannick Seurin**
Security Amplification for the Composition of Block Ciphers : Simpler Proofs and New Results
In *Selected Areas in Cryptography - SAC 2014*

[CLP14] Benoît Cogliati, Rodolphe Lampe, Jacques Patarin
The Indistinguishability of the XOR of k Permutations
In *Fast Software Encryption - FSE 2014*

Chapitre 1

Preuves de sécurité en cryptographie symétrique

1.1 La cryptographie symétrique

La cryptographie à clé secrète, ou cryptographie symétrique, est la discipline scientifique qui étudie les techniques permettant à deux personnes qui partagent un secret commun, une clé, de communiquer de façon sûre en présence d'adversaires malveillants. Il s'agit de la forme la plus ancienne de cryptographie, déjà utilisée dans l'antiquité. Parmi les exemples historiques, on peut citer le chiffre de César qui consiste en un simple décalage dans l'alphabet : le chiffré d'un texte s'obtient en décalant chaque lettre du texte clair d'un nombre fixe de lettres dans l'alphabet, toujours dans le même sens (lorsque l'on atteint une extrémité de l'alphabet, il suffit de reprendre à l'autre extrémité). Par exemple, si l'on choisit un décalage de trois lettres, A devient D, B devient E et ainsi de suite jusqu'à Z qui devient C. Le déchiffrement s'effectue de la même manière, en changeant le sens de parcours de l'alphabet. Cette technique rudimentaire présente bien évidemment de nombreux inconvénients. Citons notamment le nombre restreint de clés qui permet une recherche exhaustive très efficace. Ce chiffre a connu de nombreuses variantes au cours des siècles, comme le chiffre de Vigenère qui exploite l'algorithme de César en utilisant un décalage différent suivant la position de la lettre dans le texte. Par exemple, dans le cas d'une clé de longueur deux, les lettres en position paire dans le texte seront décalées d'une certaine valeur, tandis que celles situées en position impaire seront décalées d'une autre valeur. Cette évolution renforce d'autant plus la sécurité que la clé est longue. Lorsque la clé est aussi longue que le texte à chiffrer, et choisie aléatoirement avec une distribution uniforme, on obtient le chiffre de Vernam. En 1949, Claude Shannon a prouvé que cet algorithme est inconditionnellement sûr [Sha49]. Toutefois, sa mise en œuvre présente d'importantes difficultés techniques. En effet, il est essentiel pour la sécurité de ce chiffrement que la clé, c'est-à-dire la liste des décalages à utiliser pour chaque lettre, soit à usage unique et au moins aussi longue que le message clair : cette méthode nécessite donc de générer et de partager au préalable un grand nombre de bits de clé, ce qui est difficilement applicable en pratique, par exemple pour sécuriser les communications sur Internet.

Deux approches différentes ont été suivies pour permettre le chiffrement d'un grand volume de données à partir d'une petite quantité d'aléa, qui ont abouti à deux familles d'algorithmes :

— les algorithmes de chiffrement par flots : à partir d'une graine (une petite quantité de données aléatoires), on génère une longue suite de données pseudo-aléatoires qui sera XORée bit à bit au message, sur le modèle du chiffrement de Vernam dans le cas particulier de l'alphabet $\{0,1\}$;

— les algorithmes de chiffrement par blocs : ces algorithmes fonctionnent à l'aide d'une clé aléatoire de taille fixe, et se limitent à chiffrer des messages d'une certaine longueur ; le chiffrement d'un message de longueur arbitraire s'effectue en découpant celui-ci en blocs auxquels sera appliqué l'algorithme de chiffrement selon un processus spécifique appelé un mode d'opération.

C'est à cette dernière famille que nous allons nous intéresser dans ce manuscrit. Commençons par définir formellement la notion d'algorithme de chiffrement par blocs.

1.2 Les algorithmes de chiffrement par blocs

Soient $\mathcal{M}$ et $\mathcal{K}$ deux ensembles non-vides. On note $\mathsf{Perm}(\mathcal{M})$ l'ensemble des permutations de $\mathcal{M}$. Formellement, un algorithme de chiffrement par blocs E ayant pour espace de messages $\mathcal{M}$ et pour espace de clé $\mathcal{K}$ est une fonction du produit cartésien $\mathcal{K} \times \mathcal{M}$ dans $\mathcal{M}$ telle que, pour tout élément k de $\mathcal{K}$, la fonction

$$\begin{aligned} E(k,\cdot) \,:\, \mathcal{M} &\to \mathcal{M} \\ x &\mapsto E(k,x) \end{aligned}$$

est une permutation. À cette définition purement syntaxique s'ajoute généralement la contrainte pratique que les opérations de chiffrement et de déchiffrement soient calculables efficacement. On notera $\mathsf{BC}(\mathcal{K},\mathcal{M})$ l'ensemble de ces algorithmes de chiffrement par blocs.

L'algorithme de chiffrement le plus général possible est l'algorithme de chiffrement ayant pour espace de clé $\mathsf{Perm}(\mathcal{M})$ et qui, pour chaque permutation π de $\mathsf{Perm}(\mathcal{M})$, implémente simplement cette permutation. Cet algorithme est dit parfait puisque chiffrer des messages sous une clé uniformément aléatoire est alors équivalent à tirer les messages chiffrés aléatoirement avec une distribution uniforme dans $\mathcal{M}$, ce qui ne révèle aucune information sur le texte clair. Bien qu'il soit idéal du point de vue de la sécurité, tout comme le chiffre de Vernam, cet algorithme est inutilisable en pratique. En effet, il dispose de $|\mathcal{M}|!$ clés ; par exemple, dans le cas où $\mathcal{M} = \{0,1\}^n$ pour un entier n, cela obligerait ses utilisateurs à générer et échanger de l'ordre de $n2^n$ bits de données aléatoires, ce qui est hautement irréaliste. En effet, dans le applications courantes, on utilise des valeurs de n supérieures à 80 afin d'obtenir un espace de messages de taille suffisamment élevée pour éviter l'attaque triviale dans laquelle on constitue simplement un dictionnaire de toutes les paires clair/chiffré. Les algorithmes concrets utilisent plutôt pour clé des chaînes de caractères de petite taille, typiquement comprise entre 128 et 256 bits et cherchent à se rapprocher le plus

possible, dans un certain sens, de cet algorithme parfait tout en restant calculables et inversibles très efficacement. Une des notions permettant de comparer un algorithme de chiffrement par blocs à cet algorithme parfait, et de ce fait d'en caractériser la sécurité, est l'indistinguabilité.

1.3 La notion d'indistinguabilité

De façon informelle, un algorithme de chiffrement par blocs est jugé sûr si un attaquant est incapable de le différencier d'une permutation uniformément aléatoire lorsqu'il est équipé d'une clé choisie uniformément aléatoirement. On dit dans ce cas qu'il est pseudo-aléatoire, ou encore qu'il est indistinguable d'une permutation aléatoire.

Plus précisément, soit $E \in \mathsf{BC}(\{0,1\}^\kappa, \{0,1\}^n)$ pour deux entiers naturels quelconques κ et n. Imaginons le jeu suivant entre deux joueurs, respectivement appelés l'attaquant et le défenseur :

1. le défenseur tire à pile ou face ;
   (a) si le résultat est "pile", il tire une clé $k \in \{0,1\}^\kappa$ avec une distribution uniforme et définit $P = E_k$;
   (b) sinon, il tire aléatoirement avec une distribution uniforme une permutation π sur $\{0,1\}^n$ et définit $P = \pi$ (ce qui équivaut à équiper l'algorithme de chiffrement parfait d'une clé uniformément aléatoire) ;
2. l'attaquant choisit un message $m \in \{0,1\}^n$ et demande au défenseur son image directe ou inverse par P ;
3. l'étape précédente est répétée q fois au total ;
4. l'attaquant gagne s'il devine correctement de quel côté la pièce est tombée.

La permutation P à laquelle l'adversaire a accès est appelée un oracle et l'attaquant, modélisé comme un algorithme avec un accès à cet oracle de permutation, est appelé un distingueur. En outre, nous considérerons que les distingueurs n'effectuent jamais de requête redondante, c'est-à-dire de requête dont ils peuvent déjà connaître la réponse grâce à une requête précédente. Enfin, on supposera qu'un distingueur effectuera toujours le nombre maximal de requêtes qui lui sont allouées. Dans ces conditions, un attaquant répondant au hasard sera capable de distinguer l'algorithme de chiffrement par blocs E d'une permutation aléatoire avec une probabilité égale à $1/2$. L'indistinguabilité de E sera donc d'autant plus importante que la probabilité d'un attaquant de sortir victorieux sera proche de cette valeur.

Les distingueurs sont habituellement classifiés selon le type de requêtes qu'ils effectuent. Un distingueur qui ne peut effectuer que des requêtes directes (de chiffrement) à l'oracle est appelé un distingueur cpa (pour *chosen plaintext attack*), alors qu'il est appelé distingueur cca (pour *chosen ciphertext attack*) s'il effectue à la fois des requêtes directes et inverses (de déchiffrement). Ces deux familles ont également une variante non-adaptative, ncpa et ncca respectivement (pour *non-adaptive chosen plaintext attack* et *non-adaptive chosen ciphertext attack*), dans ce cas le distingueur doit choisir toutes ses requêtes avant de recevoir la réponse à sa première requête.

Il est maintenant temps de donner une définition formelle de ces notions. Soit $\mathcal{D}$ un distingueur. L'avantage de $\mathcal{D}$ est défini par

$$\mathbf{Adv}(\mathcal{D}) = \left|\Pr\left[k \leftarrow_{\$} \mathcal{K} : \mathcal{D}^{E_k} = 1\right] - \Pr\left[P \leftarrow_{\$} \mathsf{Perm}(\mathcal{M}) : \mathcal{D}^{P} = 1\right]\right|,$$

où, selon le type de distingueur, $\mathcal{D}$ peut effectuer des requêtes dans un seul ou dans les deux sens, de façon adaptative ou non. Cette valeur est reliée à la probabilité p qu'a $\mathcal{D}$ de distinguer l'algorithme E d'une permutation aléatoire par la formule suivante :

$$\mathbf{Adv}(\mathcal{D}) = 2\left|p - \frac{1}{2}\right|.$$

Ainsi, plus l'avantage de $\mathcal{D}$ est proche de 0, plus la probabilité p qu'a $\mathcal{D}$ de remporter le jeu précédent est proche de $1/2$ et plus l'algorithme E sera indistinguable d'une permutation aléatoire aux yeux de l'attaquant.

Pour tout entier naturel q, l'avantage de E face aux attaques de la classe ATK, où $\mathrm{ATK} \in \{(\mathrm{n})\mathrm{cpa}, (\mathrm{n})\mathrm{cca}\}$, est défini par

$$\mathbf{Adv}_E^{\mathrm{ATK}}(q) = \max_{\mathcal{D}} \mathbf{Adv}(\mathcal{D}),$$

où le maximum est pris sur l'ensemble des distingueurs $\mathcal{D}$ de type ATK effectuant au plus q requêtes à l'oracle. On dit que E est (q, ε)-résistant aux attaques ATK si $\mathbf{Adv}_E^{\mathrm{ATK}}(q) \leq \varepsilon$. Dans ce manuscrit, nous considérerons des attaquants disposant d'une puissance de calcul illimitée qui ont un nombre limité d'accès en boîte noire à un oracle : c'est-à-dire qu'ils n'ont accès à aucune autre information de la part de l'oracle que sa sortie. De plus, grâce à cette hypothèse, nous pourrons sans perte de généralité supposer que les distingueurs sont déterministes.

Cette caractérisation de la sécurité est naturelle : en effet, si un défaut dans la nature pseudo-aléatoire de l'algorithme n'est pas nécessairement fatal, l'existence d'une attaque contre un algorithme donne systématiquement lieu à un distingueur disposant d'un avantage significatif. La notion n'est toutefois pas parfaite et on peut identifier au moins deux types d'attaques qui lui échappent :

— les attaques à clés reliées : dans ce contexte, on autorise l'adversaire à influer sur la clé grâce à un ensemble de transformations fixé ; par exemple, l'adversaire a la possibilité de XORer la constante de son choix à la clé à chaque requête ;
— les attaques physiques : ces attaques utilisent le fait que les algorithmes sont implémentés et utilisés sur des dispositifs matériels, il est donc possible d'exploiter des données physiques, comme la consommation électrique par exemple, pour créer de nouvelles attaques.

Les attaques à clés reliées, bien qu'elles permettent de mieux comprendre le fonctionnement des algorithmes de chiffrement par blocs, ne sont généralement pas considérées comme étant applicables en pratique, puisqu'elle requièrent de la part de l'attaquant la capacité de modifier la clé. Dans la plupart des cas concrets, ceci n'est pas réaliste. Le cas des attaques physiques est très différent : elles sont utilisées en pratique, mais ciblent surtout l'implémentation de l'algorithme de chiffrement. C'est donc généralement à ce niveau que sont prises des mesures de protection. L'indistinguabilité reste donc à ce jour la notion la plus utilisée pour l'étude mathématique des algorithmes de chiffrement par blocs.

On peut identifier de nombreuses techniques permettant de prouver des résultats d'indistinguabilité en cryptographie symétrique. On peut notamment citer les techniques utilisant sur des étapes intermédiaires appelées « jeux », celles reposant sur la méthode probabiliste du couplage et enfin la technique des coefficients H. C'est cette dernière que nous avons principalement utilisée dans nos travaux, nous allons donc présenter brièvement son fonctionnement et l'illustrer par plusieurs exemples, de complexité croissante.

1.4 La technique des coefficients H

1.4.1 Présentation

Le contexte général de l'utilisation de cette technique est celui d'un distingueur $\mathcal{D}$ effectuant un nombre entier q de requêtes à un oracle qui peut être soit un oracle du monde réel (lorsque $\mathcal{D}$ interagit avec l'algorithme dont on veut prouver la sécurité), soit un oracle du monde idéal (lorsque $\mathcal{D}$ interagit avec une primitive idéale, par exemple une permutation aléatoire pour l'étude d'un algorithme de chiffrement par blocs). Cette interaction est enregistrée dans une transcription, qui est constituée de la liste des requêtes effectuées ainsi que des réponses reçues. Nous allons ensuite ne considérer que les transcriptions atteignables, c'est-à-dire celles dont la probabilité d'apparition dans le monde idéal est non-nulle. De façon informelle, le lemme principal de la technique des coefficients H affirme que l'avantage de $\mathcal{D}$ reste petit tant que, pour la plupart des transcriptions atteignables, les probabilités d'obtenir cette transcription dans chaque monde sont suffisamment proches. Ainsi, une preuve typique reposant sur ce résultat se divise en trois temps :

1. on partitionne l'ensemble des transcriptions atteignables en deux sous-ensembles mutuellement exclusifs : les bonnes et les mauvaises transcriptions, ces dernières étant constituées de transcriptions ayant une probabilité trop faible d'apparaître dans le monde réel, et parfois d'un petit nombre de transcriptions pour lesquelles cette probabilité est difficile à évaluer ;
2. on prouve que la probabilité qu'une mauvaise transcription apparaisse dans le monde idéal est faible ;
3. on fixe une bonne transcription et on minore le rapport entre les probabilités que cette transcription apparaisse dans le monde réel et dans le monde idéal.

Les deux dernières étapes font appel à des techniques combinatoires reposant sur l'évaluation du nombre de permutations ou de fonctions qui vérifient un certain nombre de contraintes fixées par la transcription.

1.4.2 Énoncé du résultat général

Avant d'énoncer et de démontrer le lemme central de la technique des coefficients H dans le cas le plus général, introduisons quelques notations. Lorsqu'un distingueur $\mathcal{D}$ est fixé, on notera Θ l'ensemble des transcriptions atteignables. Par la suite, on notera respectivement T_{re} et T_{id} la loi de probabilité de la transcription induite par le

monde réel, respectivement par le monde idéal. Par extension, on utilisera la même notation pour noter une variable aléatoire distribuée selon chaque loi de probabilité.

Lemme 1 ([Pat08b, CS14])**.** *Fixons un distingueur $\mathcal{D}$. Soit $\Theta = \Theta_{\text{good}} \sqcup \Theta_{\text{bad}}$ une partition de l'ensemble des transcriptions atteignables. Supposons qu'il existe ε_1 tel que, pour tout $\tau \in \Theta_{\text{good}}$, on a*

$$\frac{\Pr[T_{\text{re}} = \tau]}{\Pr[T_{\text{id}} = \tau]} \geq 1 - \varepsilon_1,$$

et qu'il existe ε_2 tel que $\Pr[T_{\text{id}} \in \Theta_{\text{bad}}] \leq \varepsilon_2$. Alors $\mathbf{Adv}(\mathcal{D}) \leq \varepsilon_1 + \varepsilon_2$.

Remarque 1. *Notons que cette définition a un sens puisque, pour toute transcription atteignable, on a $\Pr[T_{\text{id}} = \tau] > 0$.*

Avant de démontrer de ce résultat, commençons par rappeler quelques éléments sur la distance statistique entre deux distributions de probabilités.

Définition 1 (Distance statistique)**.** *Soit Ω un ensemble fini et soient μ et ν deux distributions de probabilité définies sur Ω. La* distance statistique *entre μ et ν, notée $\|\mu - \nu\|$ est définie comme :*

$$\|\mu - \nu\| = \frac{1}{2} \sum_{\omega \in \Omega} |\mu(\omega) - \nu(\omega)|.$$

Il est possible de montrer que les trois définitions suivantes sont équivalentes :

$$\|\mu - \nu\| = \max_{S \subseteq \Omega} \{\mu(S) - \nu(S)\} = \max_{S \subseteq \Omega} \{\nu(S) - \mu(S)\} = \max_{S \subseteq \Omega} \{|\mu(S) - \nu(S)|\}.$$

De plus, il est facile de montrer que

$$\|\mu - \nu\| = \sum_{\substack{\omega \in \Omega \\ \mu(\omega) \geq \nu(\omega)}} (\mu(\omega) - \nu(\omega)). \tag{1.1}$$

Nous pouvons à présent donner une preuve du lemme 1.

Démonstration. Notons $\mathcal{O}_{\text{id}}$, respectivement $\mathcal{O}_{\text{re}}$ l'oracle du monde idéal, respectivement du monde réel. On rappelle que

$$\begin{aligned} \mathbf{Adv}(\mathcal{D}) &= \left| \Pr\left[\mathcal{D}^{\mathcal{O}_{\text{id}}} = 1\right] - \Pr\left[\mathcal{D}^{\mathcal{O}_{\text{re}}} = 1\right] \right| \\ &= \left| \Pr\left[\mathcal{D}^{\mathcal{O}_{\text{id}}} = 0\right] - \Pr\left[\mathcal{D}^{\mathcal{O}_{\text{re}}} = 0\right] \right|. \end{aligned}$$

De plus, le distingueur peut être vu comme une fonction déterministe de la transcription. Si on note Θ_i l'ensemble des transcriptions atteignables pour lesquels $\mathcal{D}$ renvoie la valeur i, pour $i = 0$ ou 1, il est clair que

$$\Pr\left[\mathcal{D}^{\mathcal{O}_{\text{id}}} = i\right] = \sum_{\tau \in \Theta_i} \Pr\left[T_{\text{id}} = \tau\right] \qquad \text{et}$$

$$\Pr\left[\mathcal{D}^{\mathcal{O}_{\mathrm{re}}} = i\right] = \sum_{\tau \in \Theta_i} \Pr\left[T_{\mathrm{re}} = \tau\right]$$

pour $i = 0, 1$. Ainsi

$$\begin{aligned}\mathbf{Adv}(\mathcal{D}) &= \left| \sum_{\tau \in \Theta_1} \left(\Pr\left[T_{\mathrm{re}} = \tau\right] - \Pr\left[T_{\mathrm{id}} = \tau\right]\right) \right| \\ &\leq \sum_{\tau \in \Theta_1} |\Pr\left[T_{\mathrm{re}} = \tau\right] - \Pr\left[T_{\mathrm{id}} = \tau\right]|.\end{aligned}$$

De même,

$$\mathbf{Adv}(\mathcal{D}) \leq \sum_{\tau \in \Theta_0} |\Pr\left[T_{\mathrm{re}} = \tau\right] - \Pr\left[T_{\mathrm{id}} = \tau\right]|,$$

d'où

$$\mathbf{Adv}(\mathcal{D}) \leq \frac{1}{2} \sum_{\tau \in \Theta} |\Pr\left[T_{\mathrm{re}} = \tau\right] - \Pr\left[T_{\mathrm{id}} = \tau\right]| = \|T_{\mathrm{re}} - T_{\mathrm{id}}\|$$

puisque $\Theta = \Theta_0 \sqcup \Theta_1$ forme une partition de l'ensemble des transcriptions atteignables. De plus, on a

$$\begin{aligned}\|T_{\mathrm{re}} - T_{\mathrm{id}}\| &= \sum_{\substack{\tau \in \Theta \\ \Pr[T_{\mathrm{id}} = \tau] > \Pr[T_{\mathrm{re}} = \tau]}} \left(\Pr\left[T_{\mathrm{id}} = \tau\right] - \Pr\left[T_{\mathrm{re}} = \tau\right]\right) \\ &= \sum_{\substack{\tau \in \Theta \\ \Pr[T_{\mathrm{id}} = \tau] > \Pr[T_{\mathrm{re}} = \tau]}} \Pr\left[T_{\mathrm{id}} = \tau\right] \left(1 - \frac{\Pr\left[T_{\mathrm{re}} = \tau\right]}{\Pr\left[T_{\mathrm{id}} = \tau\right]}\right) \\ &\leq \sum_{\tau \in \Theta_{\mathrm{good}}} \Pr\left[T_{\mathrm{id}} = \tau\right] \epsilon_1 + \sum_{\tau \in \Theta_{\mathrm{bad}}} \Pr\left[T_{\mathrm{id}} = \tau\right] \\ &\leq \varepsilon_1 + \varepsilon_2.\end{aligned}$$

□

1.4.3 Le cas particulier des adversaires non-adaptatifs

Nous allons présenter ici un résultat fondamental de la technique des coefficients H [Pat90, Pat91, Pat08b], dans le cas particulier de l'avantage ncpa d'un algorithme de chiffrement par blocs. Ce cas est particulièrement intéressant puisqu'il y est possible de donner une expression exacte simple de l'avantage. Nous y ferons également appel dans le section suivante qui sera consacrée au problème de l'amplification de sécurité pour la composition d'algorithmes de chiffrement par blocs.

Fixons un espace de messages $\mathcal{M}$ quelconque et notons $M = |\mathcal{M}|$. Soit $(\mathcal{M})_q$ l'ensemble de tous les q-uplets de messages deux à deux distincts. Soit E un algorithme de chiffrement par blocs d'espace de messages $\mathcal{M}$ et d'espace de clés $\mathcal{K}_E$. Étant donné un entier $q \geq 1$ et deux q-uplets $\mathbf{x} = (x_1, \ldots, x_q) \in (\mathcal{M})_q$ et $\mathbf{y} = (y_1, \ldots, y_q) \in (\mathcal{M})_q$ d'éléments deux à deux distincts de $\mathcal{M}$, on note

$$\mathsf{p}_E(\mathbf{x}, \mathbf{y}) = \Pr\left[K \leftarrow_{\$} \mathcal{K}_E : E_K(\mathbf{x}) = \mathbf{y}\right] = \frac{|\{K \in \mathcal{K}_E : E_K(\mathbf{x}) = \mathbf{y}\}|}{|\mathcal{K}_E|},$$

où $E_K(\mathbf{x}) = \mathbf{y}$ est la notation compacte de l'évènement "$E_K(x_i) = y_i$ pour tout $1 \leq i \leq q$". De plus, on note

$$\mathsf{p}^* = \Pr\left[P \leftarrow_\$ \mathsf{Perm}(\mathcal{M}) : P(x) = y\right] = \frac{1}{M(M-1)\cdots(M-q+1)}.$$

Par extension, on notera également p^* la loi de probabilité uniforme sur $(\mathcal{M})_q$. Quand un q-uplet de messages deux à deux distincts $\mathbf{x}$ est fixé,

$$\mathsf{p}_{E,\mathbf{x}} : \mathbf{y} \mapsto \mathsf{p}_E(\mathbf{x}, \mathbf{y})$$

est la loi de probabilité, lorsque la clé K est choisie uniformément aléatoirement dans $\mathcal{K}_E$, du q-uplet de chiffrés lorsque E reçoit le q-uplet de textes clairs $\mathbf{x}$. De même, lorsqu'un q-uplet de messages deux à deux distincts $\mathbf{y}$ est fixé,

$$\mathsf{p}_{E^{-1},\mathbf{y}} : \mathbf{x} \mapsto \mathsf{p}_E(\mathbf{x}, \mathbf{y})$$

est la loi de probabilité du q-uplet de texte clairs lorsque E^{-1} reçoit en entrée le q-uplet de chiffrés $\mathbf{y}$.

On a alors le résultat suivant, dont on donnera la preuve par souci d'exhaustivité.

Lemme 2 (Sécurité ncpa [Pat08a])**.** *Soit E un algorithme de chiffrement par bloc d'espace de messages $\mathcal{M}$. Alors*

$$\mathbf{Adv}_E^{\mathrm{ncpa}}(q) = \max_{\mathbf{x} \in (\mathcal{M})_q} \|\mathsf{p}_{E,\mathbf{x}} - \mathsf{p}^*\|.$$

Démonstration. Fixons un distingueur ncpa quelconque $\mathcal{D}$. Puisque nous ne considérons que les distingueurs déterministes, $\mathcal{D}$ est complètement caractérisé par son q-uplet de requêtes deux à deux distinctes $\mathbf{x} = (x_1 \ldots, x_q)$ et par sa fonction de décision $\phi_{\mathcal{D}} : (\mathcal{M})_q \to \{0, 1\}$, où $\phi_{\mathcal{D}}(\mathbf{y})$ est la sortie de $\mathcal{D}$ lorsqu'il reçoit le q-uplet $\mathbf{y} = (y_1, \ldots, y_q)$ en réponse à ses requêtes. Par définition de l'avantage,

$$\begin{aligned}
\mathbf{Adv}(\mathcal{D}) &= \Bigg| \sum_{\mathbf{y} \in (\mathcal{M})_q : \phi_{\mathcal{D}}(\mathbf{y}) = 1} \Pr\left[K \leftarrow_\$ \mathcal{K} : E_K(\mathbf{x}) = \mathbf{y}\right] \\
&\qquad - \sum_{\mathbf{y} \in (\mathcal{M})_q : \phi_{\mathcal{D}}(\mathbf{y}) = 1} \Pr\left[P \leftarrow_\$ \mathsf{Perm}(\mathcal{M}) : P(\mathbf{x}) = \mathbf{y}\right] \Bigg| \\
&= \Bigg| \sum_{\mathbf{y} \in (\mathcal{M})_q : \phi_{\mathcal{D}}(\mathbf{y}) = 1} (\mathsf{p}_{E,\mathbf{x}}(\mathbf{y}) - \mathsf{p}^*) \Bigg| \\
&\leq \|\mathsf{p}_{E,\mathbf{x}} - \mathsf{p}^*\|.
\end{aligned}$$

En prenant le maximum sur les valeurs $\mathbf{x} \in (\mathcal{M})_q$, on obtient

$$\mathbf{Adv}_E^{\mathrm{ncpa}}(q) \leq \max_{\mathbf{x} \in (\mathcal{M})_q} \|\mathsf{p}_{E,\mathbf{x}} - \mathsf{p}^*\|.$$

Afin de prouver l'égalité de ces deux quantités, considérons le distingueur $\mathcal{D}$ dont les requêtes correspondent au q-uplet de messages deux à deux distincts $\mathbf{x}$ pour lequel $\|\mathsf{p}_{E,\mathbf{x}} - \mathsf{p}^*\|$ est maximal, et renvoie 1 si et seulement si la réponse $\mathbf{y}$ de l'oracle de

permutation vérifie l'inégalité $\mathsf{p}_E(\mathbf{x}, \mathbf{y}) \geq \mathsf{p}^*$. Notons Φ sa fonction de décision. Nous allons prouver que l'avantage de ce distingueur est exactement $\|\mathsf{p}_{E,\mathbf{x}} - \mathsf{p}^*\|$. Pour ce faire, nous allons utiliser l'égalité (1.1) :

$$\begin{aligned}\|\mathsf{p}_{E,\mathbf{x}} - \mathsf{p}^*\| &= \sum_{\substack{\mathbf{y}\in(\mathcal{M})_q \\ \mathsf{p}_{E,\mathbf{x}}(\mathbf{y})\geq\mathsf{p}^*}} (\mathsf{p}_{E,\mathbf{x}}(\mathbf{y}) - \mathsf{p}^*) \\ &= \left| \sum_{\substack{\mathbf{y}\in(\mathcal{M})_q \\ \Phi(\mathbf{y})=1}} (\mathsf{p}_{E,\mathbf{x}}(\mathbf{y}) - \mathsf{p}^*) \right| \\ &= \mathbf{Adv}(\mathcal{D}),\end{aligned}$$

d'où l'égalité recherchée. □

1.5 Illustration 1 : l'amplification de sécurité

1.5.1 Le problème de l'amplification de sécurité pour la composition d'algorithmes de chiffrement par blocs

Le problème de l'amplification de sécurité est de déterminer si combiner plusieurs algorithmes de chiffrement par blocs $E_1, \ldots, E_n$ peut donner un algorithme de chiffrement par blocs F disposant de meilleures garanties de sécurité que chacune de ses composantes. Ce problème s'étend naturellement aux autres primitives cryptographiques telles que les fonctions ou les générateurs pseudo-aléatoires, mais dans ce manuscrit nous nous concentrerons sur les permutations pseudo-aléatoires, c'est-à-dire les algorithmes de chiffrement par blocs. Le sujet de l'amplification, dans le cas des algorithmes de chiffrement par blocs, a fait l'objet de nombreux articles. On peut notamment mentionner une longue suite de travaux qui ont considéré la sécurité prouvable du chiffrement en cascade *dans le modèle de l'algorithme de chiffrement idéal* [BR06a, GM09, Lee13], qui est assez orthogonal à notre cadre : travailler dans ce modèle idéalisé équivaut en un certain sens à borner la connaissance qu'a l'adversaire de l'algorithme de chiffrement par bloc sous-jacent (puisqu'il ne peut y faire qu'un nombre limité de requêtes), alors que nous considérons des adversaires dont la puissance de calcul n'est pas bornée, dans le modèle standard, non idéalisé. En particulier, l'adversaire a une connaissance totale du ou des algorithmes de chiffrement par blocs sous-jacents, et peut, par exemple, les représenter comme un tableau géant. Dans ce contexte, on peut envisager deux types d'amélioration des garanties de sécurité d'un algorithme de chiffrement par blocs :

- soit F est plus difficile à distinguer d'une permutation aléatoire pour une classe de distingueurs donnée que chacune de ses composantes, ce que nous appellerons *ε-amplification* ;
- soit F peut résister à une classe plus vaste d'adversaire que chacune de ses composantes, ce que nous appellerons *amplification de classe*.

Commençons par donner un exemple de résultat appartenant à chacune de ces catégories.

L'exemple classique de résultat de type ε-amplification exprime le fait que composer deux algorithmes de chiffrement par blocs F et G qui sont respectivement (q, ε_F)- et (q, ε_G)-résistants aux attaques ncpa (respectivement cpa) donne un algorithme de chiffrement par blocs qui est $(q, 2\varepsilon_F\varepsilon_G)$-résistant aux attaques ncpa (respectivement cpa). Ainsi, quand $\varepsilon_F, \varepsilon_G < 1/2$, le nouvel algorithme est strictement plus sûr que chacune de ses composantes. Ce résultat a été démontré dans le cas d'adversaires dont la puissance de calcul n'est pas bornée par Vaudenay (voir [Vau98] pour la preuve dans le cas non-adaptatif et [Vau99] pour celle dans le cas adaptatif) en utilisant la *théorie de la décorrélation* [Vau03]. Maurer et Tessaro ont par la suite démontré un résultat analogue dans le cas où la puissance de calcul des adversaires est bornée [MT09].

Pour la seconde catégorie de résultats, l'exemple le plus notoire est celui que l'on appelle le théorème « deux faibles font un fort » (« two weak make one strong » en anglais, que nous abrégerons en *2W1S*), qui affirme que si F et G sont respectivement (q, ε_F)- et (q, ε_G)-résistants aux attaques ncpa, alors la composition $G^{-1} \circ F$ est $(q, \varepsilon_F + \varepsilon_G)$-résistante aux attaques cca, ce résultat étant optimal en général. Notons qu'ici, l'algorithme de chiffrement par blocs obtenu peut résister à des attaques bien plus puissantes que chaque composante F et G. Ce résultat a tout d'abord été prouvé à un terme logarithmique près par Maurer et Pietrzak [MP04], tandis que la version optimale a ensuite été démontrée par Maurer, Pietrzak, and Renner [MPR07] dans le cadre des systèmes aléatoires [Mau02]. Il est important de noter que ce résultat n'est vrai que lorsque l'attaquant dispose d'une puissance de calcul non bornée. Dans le modèle calculatoire, la composition d'algorithmes de chiffrement par blocs sûrs face à des adversaires non-adaptatifs ne donne pas, en général, un algorithme sûr face à des adversaires adaptatifs [Mye04, Pie05a], bien qu'il existe des résultats partiellement positifs à ce sujet [LR86, Pie06].

Ce théorème *2W1S* est un résultat important qui a servi de base à un certain nombre d'articles, notamment des travaux utilisant des preuves de sécurité reposant sur la technique du couplage [MRS09, HR10, LPS12, LS14]). Toutefois sa preuve repose sur une technique au formalisme très lourd, qui augmente le risque d'erreur. En effet, Jetchev *et al.* [JÖS12] ont mis en évidence le fait qu'un lemme essentiel de la théorie des systèmes aléatoires, le théorème 2 de [Mau02] (notamment utilisé par la preuve du théorème 2W1S [MPR07]), a été formulé de façon inexacte et que, de plus, la seule preuve connue de ce lemme [Pie05b] comportait une faille, qu'ils ont corrigée. Finalement, on peut considérer que la preuve de ce théorème important est obtenue en combinant trois articles différents [Mau02, MPR07, JÖS12]. De plus, à cause des difficultés conceptuelles liées aux techniques utilisées, il paraît difficile de généraliser ce résultat à une cascade plus longue. Considérons le problème suivant : étant donné trois algorithmes de chiffrement par blocs (ou plus) qui sont simplement sûrs face à des adversaires non-adaptatifs, peut-on avoir les deux types d'amplification de sécurité simultanément de façon optimale ? Notons que ceci requiert au moins une cascade de longueur trois puisque, à partir de deux algorithmes de chiffrement par blocs résistants aux attaques ncpa F et G, on peut soit obtenir l'ε-amplification en considérant $G \circ F$, soit l'amplification de classe en considérant $G^{-1} \circ F$, mais pas le deux. Concentrons-nous sur la composition d'un même algorithme et considérons un

algorithme de chiffrement par blocs E tel qu'à la fois E et E^{-1} soient (q, ε)-résistants aux attaques ncpa. Un grand nombre d'algorithmes de chiffrement par blocs ont une sécurité prouvée similaire dans le sens direct et dans le sens inverse grâce à une structure proche d'une involution, comme par exemple les schémas de Feistel symétriques. Que peut-on dire de la sécurité cca de la cascade E^n de longueur n ? En utilisant les meilleurs résultats connus, une réponse directe (en supposant n pair) peut être obtenue en appliquant récursivement le théorème d'ε-amplification pour les algorithmes de chiffrement par blocs résistants aux attaques ncpa à chaque moitié de la cascade, ce qui donne

$$\mathbf{Adv}^{\mathrm{ncpa}}_{E^{n/2}}(q) \leq 2^{\frac{n}{2}-1}\varepsilon^{\frac{n}{2}} \qquad \text{et} \qquad \mathbf{Adv}^{\mathrm{ncpa}}_{(E^{n/2})^{-1}}(q) \leq 2^{\frac{n}{2}-1}\varepsilon^{\frac{n}{2}},$$

puis le théorème 2W1S pour obtenir

$$\mathbf{Adv}^{\mathrm{cca}}_{E^n}(q) \leq \mathbf{Adv}^{\mathrm{ncpa}}_{E^{n/2}}(q) + \mathbf{Adv}^{\mathrm{ncpa}}_{(E^{n/2})^{-1}}(q) \leq (2\varepsilon)^{\frac{n}{2}}.$$

Si n est impair, en divisant E^n en deux parties déséquilibrées, un raisonnement similaire donne

$$\mathbf{Adv}^{\mathrm{cca}}_{E^n}(q) \leq \mathbf{Adv}^{\mathrm{ncpa}}_{E^{(n+1)/2}}(q) + \mathbf{Adv}^{\mathrm{ncpa}}_{(E^{(n-1)/2})^{-1}}(q) \leq 2^{\frac{n-1}{2}}\varepsilon^{\frac{n+1}{2}} + 2^{\frac{n-3}{2}}\varepsilon^{\frac{n-1}{2}}.$$

En particulier, pour $n = 3$, le meilleur résultat que l'on puisse prouver est que

$$\mathbf{Adv}^{\mathrm{cca}}_{E^3} \leq \varepsilon + 2\varepsilon^2.$$

Ainsi, on obtient l'ε-amplification à partir $n \geq 4$, en supposant $\varepsilon < 1/4$.

1.5.2 Énoncé du résultat et discussion

Dans cette sous-section, nous allons prouver grâce à la technique des coefficients H que la sécurité cca de E^n est en fait nettement meilleure, à savoir

$$\mathbf{Adv}^{\mathrm{cca}}_{E^n}(q) \leq (2\varepsilon)^{n-1}.$$

Ainsi, lorsque $n \geq 3$, ce résultat offre à la fois l'ε-amplification *et* l'amplification de classe dès lors que

$$\varepsilon < \frac{1}{2 \cdot 2^{1/(n-2)}}$$

en particulier, dès que $\varepsilon < 1/4$ pour tout $n \geq 3$. Nous allons en fait prouver le théorème suivant qui est plus général et a en particulier le corollaire intéressant suivant : soient E, F et G trois algorithmes de chiffrement par blocs tels que E, F, F^{-1} and G^{-1} sont (q, ε)-résistants aux attaques ncpa. Alors la composition $G \circ F \circ E$ est $(q, 4\varepsilon^2)$-résistante aux attaques cca.

Théorème 1 ([CPS14])**.** *Soient $E_1, \ldots, E_n$ n algorithmes de chiffrement par blocs partageant le même espace de messages $\mathcal{M}$. Pour tout entier naturel q, on a*

$$\mathbf{Adv}^{\mathrm{cca}}_{E_n \circ \cdots \circ E_1}(q) \leq 2^{n-1} \max_{1 \leq i \leq n} \left(\prod_{1 \leq j \leq i-1} \mathbf{Adv}^{\mathrm{ncpa}}_{E_j}(q) \times \prod_{i+1 \leq j \leq n} \mathbf{Adv}^{\mathrm{ncpa}}_{E_j^{-1}}(q) \right).$$

Remarque 2. *La borne du théorème 1 n'est déjà pas optimale en général pour $n = 2$. En effet il est facile de vérifier que le théorème 2W1S offre une meilleure borne, notamment lorsque E_1 et E_2^{-1} ont des niveaux de sécurité* ncpa *différents. Dans l'annexe A.3, nous présentons une version plus précise de ce résultat dans le cas particulier où $n = 3$.*

Corollaire 1 ([CPS14])**.** *Soient $E_1, \ldots, E_n$ n algorithmes de chiffrement par blocs partageant le même espace de messages $\mathcal{M}$. Fixons un entier $q \geq 1$. Pour $i = 1, \ldots, n$, soit $\varepsilon_i = \max\{\mathbf{Adv}^{\mathrm{ncpa}}_{E_i}(q), \mathbf{Adv}^{\mathrm{ncpa}}_{E_i^{-1}}(q)\}$. Alors on a*

$$\mathbf{Adv}^{\mathrm{cca}}_{E_n \circ \cdots \circ E_1}(q) \leq 2^{n-1} \max_{1 \leq i \leq n} \prod_{\substack{1 \leq j \leq n \\ j \neq i}} \varepsilon_j.$$

Remarque 3. *Il n'est pas difficile de voir que le corollaire 1 reste vrai lorsque $\varepsilon_1 = \mathbf{Adv}^{\mathrm{ncpa}}_{E_1}(q)$ et $\varepsilon_n = \mathbf{Adv}^{\mathrm{ncpa}}_{E_n^{-1}}$, c'est-à-dire que E_1 et E_n n'ont besoin d'être sûrs que dans une seule direction. Seules les composantes « internes » $E_2, \ldots, E_{n-1}$ ont besoin d'être sûres dans les deux directions.*

Dans le cas où l'on compose un algorithme de chiffrement par blocs avec lui-même, on a le résultat suivant.

Corollaire 2 ([CPS14])**.** *Soient E un algorithme de chiffrement par blocs et $q \geq 1$. Notons*

$$\varepsilon = \max\{\mathbf{Adv}^{\mathrm{ncpa}}_{E}(q), \mathbf{Adv}^{\mathrm{ncpa}}_{E^{-1}}(q)\}.$$

Alors, pour tout entier $n \geq 1$,

$$\mathbf{Adv}^{\mathrm{cca}}_{E^n}(q) \leq (2\varepsilon)^{n-1}.$$

Remarque 4. *Le prérequis nécessaire à l'application du corollaire 2, à savoir qu'à la fois E et E^{-1} soient (q, ε)-résistants aux attaques* ncpa*, peut sembler beaucoup plus fort que la simple supposition que l'algorithme E soit (q, ε)-résistant aux attaques* ncpa*. Cependant, les schémas utilisés dans les algorithmes de chiffrement par blocs sont souvent des involutions ou proches d'involutions (par exemple les schémas de Feistel symétriques). Alors il n'est nécessaire de déterminer qu'*une seule *de ces bornes. Il existe toutefois des structures d'algorithmes de chiffrement par blocs tels que la sécurité* ncpa *de E^{-1} est beaucoup moins bonne que celle de E, un exemple important étant les schémas de Feistel généralisés de type 1 [ZMI89, MV00], sur lesquels est fondé par exemple l'algorithme* CAST-256*.*

1.5.3 Notations et description des transcriptions

Étant donné deux algorithmes de chiffrement par blocs E et F disposant du même espace de messages $\mathcal{M}$ et ayant pour espaces de clés respectifs $\mathcal{K}_E$ et $\mathcal{K}_F$, on note $F \circ E$ l'algorithme de chiffrement par blocs ayant pour espace de messages $\mathcal{M}$ et pour ensemble de clés $\mathcal{K}_E \times \mathcal{K}_F$ défini par

$$F \circ E_{(K_E, K_F)}(x) = F_{K_F}(E_{K_E}(x)).$$

Nous appellerons $F \circ E$ la *composition* ou la *cascade* de E et F de façon interchangeable. Cette définition s'étend naturellement à la composition de $n > 2$ algorithmes de chiffrement par blocs. On note E^n l'algorithme obtenu en composant n fois E avec lui-même, sous des clés indépendantes.

Fixons deux entiers naturels n et q, ainsi que n algorithmes de chiffrement par blocs $E_1, \ldots, E_n$ ayant le même espace de messages $\mathcal{M}$ et pour espaces de clés respectifs $\mathcal{K}_1, \ldots, \mathcal{K}_n$. Soit $E = E_n \circ \ldots \circ E_1$ la cascade de $E_1, \ldots, E_n$. Fixons également un distingueur $\mathcal{D}$ effectuant q requêtes et cherchant à distinguer E d'une permutation de $\mathcal{M}$ choisie uniformément aléatoirement. Rappelons que, dans ce contexte, $\mathcal{D}$ interagit avec un unique oracle implémentant une permutation :

— dans le cas réel, il s'agit de l'algorithme de chiffrement par blocs E, utilisé avec une clé choisie uniformément aléatoirement ;
— dans le cas idéal, il s'agit d'une permutation uniformément aléatoire.

On notera P cet oracle de façon générique. La transcription $\mathcal{Q}$ des requêtes de l'interaction de $\mathcal{D}$ avec P est une liste de paires de messages constituée de la façon suivante : si $\mathcal{D}$ effectue une requête directe x, et reçoit pour réponse y, ou si $\mathcal{D}$ effectue la requête inverse y, et reçoit pour réponse x, alors la paire $(x, y) \in \mathcal{M}^2$ est ajoutée à $\mathcal{Q}$. Pour conclure, une transcription τ est donc simplement la transcription $\mathcal{Q}$ des requêtes de l'adversaire. Par la suite, on notera T_{re}, respectivement T_{id}, la distribution de probabilité de la transcription τ induite par le monde réel, respectivement le monde idéal (notons que ces deux distributions de probabilité dépendent du distingueur). Par extension, on utilise la même notation pour noter une variable aléatoire distribuée selon chaque distribution.

Suivant la méthode générale de la technique des coefficients H, il faut tout d'abord définir un sous-ensemble Θ_{bad} des transcriptions atteignables constitué de mauvaises transcriptions et majorer la probabilité que T_{id} soit dans l'ensemble Θ_{bad}. Dans cet exemple, nous allons simplement choisir $\Theta_{\text{bad}} = \varnothing$, et nous appliquerons le lemme 1 avec $\varepsilon_2 = 0$.

1.5.4 Étude des bonnes transcriptions

Fixons à présent une bonne transcription $\tau = ((x_1, y_1), \ldots, (x_q, y_q))$. Nous devons à présent étudier le rapport $\Pr[T_{\text{re}} = \tau] / \Pr[T_{\text{id}} = \tau]$. Notons $\mathbf{x} = (x_1, \ldots, x_q)$ et $\mathbf{y} = (y_1, \ldots, y_q)$. Puisque τ est une transcription atteignable, on a $\mathbf{x} \in (\mathcal{M})_q$ et $\mathbf{y} \in (\mathcal{M})_q$. Ainsi, en utilisant les notations définies dans la section 1.4, on a

$$\Pr[T_{\text{id}} = \tau] = \mathsf{p}^*(\mathbf{x}, \mathbf{y}) \tag{1.2}$$

et

$$\Pr[T_{\text{re}} = \tau] = \mathsf{p}_E(\mathbf{x}, \mathbf{y}). \tag{1.3}$$

Les lemmes 3 et 4 vont nous permettre d'exprimer exactement $\mathsf{p}_E(\mathbf{x}, \mathbf{y})$ en fonction de p^* et des distributions de probabilité p_{E_i} pour $i = 1, \ldots, n$.

Lemme 3. *Soient F et G deux algorithmes de chiffrement par blocs partageant le même espace de messages $\mathcal{M}$ et ayant respectivement $\mathcal{K}_F$ et $\mathcal{K}_G$ pour espace de clés.*

Alors, pour tout q-uplet $\mathbf{u}$ et tout q-uplet $\mathbf{v}$ d'éléments deux à deux distincts de $\mathcal{M}$, on a

$$\mathsf{p}_{G\circ F}(\mathbf{u},\mathbf{v}) = \mathsf{p}^* + \sum_{\mathbf{z}\in(\mathcal{M})_q} (\mathsf{p}_F(\mathbf{u},\mathbf{z}) - \mathsf{p}^*)(\mathsf{p}_G(\mathbf{z},\mathbf{v}) - \mathsf{p}^*). \tag{1.4}$$

Démonstration. Soient $\mathbf{u} = (u_1,\dots,u_q) \in (\mathcal{M})_q$ et $\mathbf{v} = (v_1,\dots,v_q) \in (\mathcal{M})_q$. Remarquons tout d'abord que

$$\sum_{\mathbf{z}\in(\mathcal{M})_q} (\mathsf{p}_F(\mathbf{u},\mathbf{z}) - p^*) = \sum_{\mathbf{z}\in(\mathcal{M})_q} (\mathsf{p}_F(\mathbf{z},\mathbf{v}) - p^*) = 0. \tag{1.5}$$

On a

$$\begin{aligned}
\mathsf{p}_{G\circ F}(\mathbf{u},\mathbf{v}) &= \sum_{\mathbf{z}\in(\mathcal{M})_q} \mathsf{p}_F(\mathbf{u},\mathbf{z})\mathsf{p}_G(\mathbf{z},\mathbf{v}) \\
&= \sum_{\mathbf{z}} (\mathsf{p}_F(\mathbf{u},\mathbf{z}) - \mathsf{p}^* + \mathsf{p}^*)(\mathsf{p}_G(\mathbf{z},\mathbf{v}) - \mathsf{p}^* + \mathsf{p}^*) \\
&= \sum_{\mathbf{z}} (\mathsf{p}_F(\mathbf{u},\mathbf{z}) - \mathsf{p}^*)(\mathsf{p}_G(\mathbf{z},\mathbf{v}) - \mathsf{p}^*) + \mathsf{p}^* \underbrace{\sum_{\mathbf{z}} (\mathsf{p}_F(\mathbf{u},\mathbf{z}) - \mathsf{p}^*)}_{=0 \text{ par } (1.5)} \\
&\quad + \mathsf{p}^* \underbrace{\sum_{\mathbf{z}} (\mathsf{p}_G(\mathbf{z},\mathbf{v}) - \mathsf{p}^*)}_{=0 \text{ par } (1.5)} + \underbrace{\sum_{\mathbf{z}} (\mathsf{p}^*)^2}_{=\mathsf{p}^*} \\
&= \mathsf{p}^* + \sum_{\mathbf{z}} (\mathsf{p}_F(\mathbf{u},\mathbf{z}) - \mathsf{p}^*)(\mathsf{p}_G(\mathbf{z},\mathbf{v}) - \mathsf{p}^*).
\end{aligned}$$

□

Le lemme 3 est une brique essentielle de notre preuve. Il permet en effet d'exprimer la probabilité de transition pour la composition de deux algorithmes de chiffrement par blocs comme une covariance. Nous avons constaté que cette égalité peut être utilisée pour prouver d'autres résultats. Nous en présenterons deux dans l'annexe A :

- une nouvelle preuve élémentaire du théorème d'amplification de sécurité face aux adversaires ncpa pour la composition d'algorithmes de chiffrement par blocs (initialement démontré par Vaudenay dans [Vau98]) dans l'annexe A.1 ;
- un nouveau résultat de sécurité face aux attaques à clair connu dans l'annexe A.2.

Lemme 4. *On a*

$$\mathsf{p}_{E_n\circ\cdots\circ E_1}(\mathbf{x},\mathbf{y}) = \mathsf{p}^* + \sum_{\mathbf{x}_1,\dots,\mathbf{x}_{n-1}\in(\mathcal{M})_q} \left(\prod_{i=1}^{n} (\mathsf{p}_{E_i}(\mathbf{x}_{i-1},\mathbf{x}_i) - \mathsf{p}^*) \right) \tag{1.6}$$

où $\mathbf{x}_0 := \mathbf{x}$ et $\mathbf{x}_n := \mathbf{y}$.

Démonstration. La preuve se fait par récurrence. Pour tout entier $i \geq 1$, soit (H_i) la proposition suivante : pour tout $j \in \{1,\dots,i\}$, pour tout q-uplet $\mathbf{x}_j$ d'éléments deux à deux distincts de $\mathcal{M}$, on a

$$\mathsf{p}_{E_j\circ\cdots\circ E_1}(\mathbf{x}_0,\mathbf{x}_j) = \mathsf{p}^* + \sum_{\mathbf{x}_1,\dots,\mathbf{x}_{j-1}\in(\mathcal{M})_q} \left(\prod_{i=1}^{j} (\mathsf{p}_{E_i}(\mathbf{x}_{i-1},\mathbf{x}_i) - \mathsf{p}^*) \right).$$

Le lemme 3 correspond à (H_2).

Supposons que (H_k) est vérifié pour un entier $k \geq 2$. Soit $\mathbf{x}_{k+1} \in (\mathcal{M})_q$. Alors ($H_i$) est vérifié pour $i = 1, \ldots, k$. Il reste à montrer que (H_{k+1}) est vraie. On a

$$\begin{aligned}
\mathsf{p}_{E_{k+1}\circ\cdots\circ E_1}(\mathbf{x}_0, \mathbf{x}_{k+1}) &= \mathsf{p}^* + \sum_{\mathbf{x}_1 \in (\mathcal{M})_q} (\mathsf{p}_{E_1}(\mathbf{x}_0, \mathbf{x}_1) - \mathsf{p}^*) \\
&\qquad \times (\mathsf{p}_{E_{k+1}\circ\cdots\circ E_2}(\mathbf{x}_1, \mathbf{x}_{k+1}) - \mathsf{p}^*) \qquad (H_2) \\
&= \mathsf{p}^* + \sum_{\mathbf{x}_1 \in (\mathcal{M})_q} (\mathsf{p}_{E_1}(\mathbf{x}_0, \mathbf{x}_1) - \mathsf{p}^*) \\
&\qquad \times \sum_{\substack{\mathbf{x}_2,\ldots,\mathbf{x}_k \\ \in (\mathcal{M})_q}} \prod_{i=2}^{k+1} (\mathsf{p}_{E_i}(\mathbf{x}_{i-1}, \mathbf{x}_i) - \mathsf{p}^*) \qquad (H_k)
\end{aligned}$$

d'où on déduit le résultat. □

Nous devons maintenant étudier la somme apparaissant dans le membre de droite de (1.6) en divisant la somme selon le signe de chaque terme du produit. Dans le but de rendre les notations plus compactes, pour tout uplet $(\mathbf{t}_0, \ldots, \mathbf{t}_n) \in ((\mathcal{M})_q)^{n+1}$ et pour tout $i \in \{1, \ldots, n\}$ on note :

— $C_{0,i}$ l'inégalité $\mathsf{p}_{E_i}(\mathbf{t}_{i-1}, \mathbf{t}_i) - \mathsf{p}* > 0$ et

— $C_{1,i}$ l'inégalité $\mathsf{p}_{E_i}(\mathbf{t}_{i-1}, \mathbf{t}_i) - \mathsf{p}* < 0$.

Alors chaque terme de la somme peut être paramétré par un $n-$uplet $\mathbf{k} = (k_1, \ldots, k_n)$ d'entiers naturels dans $\{0, 1\}$, le produit étant positif si et seulement si $k_1 + \ldots + k_n \equiv 0 \bmod 2$. Bien évidemment, les cas qui doivent être traités avec attention sont ceux pour lesquels le produit est négatif (c'est-à-dire que $k_1 + \ldots + k_n \equiv 1 \bmod 2$). C'est ce qui est fait dans le lemme suivant.

Lemme 5. *Soit* $\mathbf{k} = (k_1, \ldots, k_n) \in \{0, 1\}^n$ *tel que* $k_1 + \ldots + k_n \equiv 1 \bmod 2$. *Pour toute paire de q-uplets fixés* $\mathbf{t}_0, \mathbf{t}_n$ *in* $(\mathcal{M})_q$, *notons*

$$A_{\mathbf{k}}(\mathbf{t}_0, \mathbf{t}_n) := \{\mathbf{t} = (\mathbf{t}_1, \ldots, \mathbf{t}_{n-1}) \in ((\mathcal{M})_q)^{n-1} \,|\, \forall i \in \{1, \ldots, n\}, C_{k_i,i} \text{ est vérifié}\}.$$

Alors

$$\begin{aligned}
&\sum_{\mathbf{t} \in A_{\mathbf{k}}(\mathbf{t}_0, \mathbf{t}_n)} \prod_{1 \leq i \leq n} (\mathsf{p}_{E_i}(\mathbf{t}_{i-1}, \mathbf{t}_i) - \mathsf{p}^*) \\
&\qquad \geq -\mathsf{p}^* \max_{1 \leq i \leq n} \left(\prod_{1 \leq j \leq i-1} \mathbf{Adv}^{\mathrm{ncpa}}_{E_j}(q) \times \prod_{i+1 \leq j \leq n} \mathbf{Adv}^{\mathrm{ncpa}}_{E_j^{-1}}(q) \right).
\end{aligned}$$

Démonstration. Puisque $k_1 + \ldots + k_n \equiv 1 \bmod 2$, on peut trouver un indice j tel que $k_j = 1$, c'est-à-dire, par définition, $\mathsf{p}_{E_j}(\mathbf{t}_{j-1}, \mathbf{t}_j) - \mathsf{p}^* < 0$. Alors, on a

$$\sum_{\mathbf{t} \in A_{\mathbf{k}}(\mathbf{t}_0, \mathbf{t}_n)} \prod_{1 \leq i \leq n} (\mathsf{p}_{E_i}(\mathbf{t}_{i-1}, \mathbf{t}_i) - \mathsf{p}^*) \geq -\mathsf{p}^* \sum_{\mathbf{t} \in A_{\mathbf{k}}(\mathbf{t}_0, \mathbf{t}_n)} \prod_{\substack{1 \leq i \leq n \\ i \neq j}} (\mathsf{p}_{E_i}(\mathbf{t}_{i-1}, \mathbf{t}_i) - \mathsf{p}^*).$$

Chaque terme de la somme apparaissant dans le membre de droite est positif puisque il y a un nombre pair de termes négatifs dans chaque produit. Ainsi,

$$\sum_{\mathbf{t} \in A_{\mathbf{k}}(\mathbf{t}_0, \mathbf{t}_n)} \prod_{1 \leq i \leq n} (\mathsf{p}_{E_i}(\mathbf{t}_{i-1}, \mathbf{t}_i) - \mathsf{p}^*) \geq -\mathsf{p}^* \sum_{\mathbf{t} \in A_{\mathbf{k}}(\mathbf{t}_0, \mathbf{t}_n)} \prod_{\substack{1 \leq i \leq n \\ i \neq j}} |\mathsf{p}_{E_i}(\mathbf{t}_{i-1}, \mathbf{t}_i) - \mathsf{p}^*|.$$

Soient

$$B := \{(\mathbf{t}_1, \ldots, \mathbf{t}_{j-1}) \in ((\mathcal{M})_q)^{j-1} \,|\, \forall i \in \{1, \ldots, j-1\}, C_{k_i,i} \text{ est vrai}\} \text{ et}$$
$$C := \{(\mathbf{t}_j, \ldots, \mathbf{t}_{n-1}) \in ((\mathcal{M})_q)^{n-j} \,|\, \forall i \in \{j+1, \ldots, n\}, C_{k_i,i} \text{ est vrai}\}.$$

On a $A_{\mathbf{k}}(\mathbf{t}_0, \mathbf{t}_n) \subseteq B \times C$ puisque la seule différence entre ces ensembles est le fait que, dans $B \times C$ on a ignoré la condition $C_{k_j,j}$, c'est-à-dire que l'inégalité $\mathsf{p}_{E_j}(\mathbf{t}_{j-1}, \mathbf{t}_j) < \mathsf{p}^*$ est vérifiée. Ainsi,

$$\begin{aligned}
\sum_{t \in A_k(t_0,t_n)} \prod_{1 \le i \le n} (\mathsf{p}_{E_i}(t_{i-1}, t_i) - \mathsf{p}^*) &\ge -\mathsf{p}^* \sum_{t \in B \times C} \prod_{\substack{1 \le i \le n \\ i \ne j}} |\mathsf{p}_{E_i}(t_{i-1}, t_i) - \mathsf{p}^*| \\
&\ge -\mathsf{p}^* \underbrace{\left(\sum_{(t_1,\ldots,t_{j-1}) \in B} \prod_{1 \le i \le j-1} |\mathsf{p}_{E_i}(t_{i-1}, t_i) - \mathsf{p}^*| \right)}_{S_1} \\
&\quad \times \underbrace{\left(\sum_{(t_j,\ldots,t_{n-1}) \in C} \prod_{j+1 \le i \le n} |\mathsf{p}_{E_i}(t_{i-1}, t_i) - \mathsf{p}^*| \right)}_{S_2}.
\end{aligned}$$

Nous allons étudier ces deux sommes S_1 et S_2 indépendamment. Pour S_1, on a

$$\begin{aligned}
S_1 &= \sum_{\substack{\mathbf{t}_1 \in (\mathcal{M})_q : \\ C_{k_1,1}}} |\mathsf{p}_{E_1}(\mathbf{t}_0, \mathbf{t}_1) - \mathsf{p}^*| \ldots \sum_{\substack{\mathbf{t}_{j-1} \in (\mathcal{M})_q : \\ C_{k_{j-1},j-1}}} |\mathsf{p}_{E_{j-1}}(\mathbf{t}_{j-2}, \mathbf{t}_{j-1}) - \mathsf{p}^*| \\
&\le \sum_{\substack{\mathbf{t}_1 \in (\mathcal{M})_q : \\ C_{k_1,1}}} |\mathsf{p}_{E_1}(\mathbf{t}_0, \mathbf{t}_1) - \mathsf{p}^*| \ldots \sum_{\substack{\mathbf{t}_{j-2} \in (\mathcal{M})_q : \\ C_{k_{j-2},j-2}}} |\mathsf{p}_{E_{j-2}}(\mathbf{t}_{j-3}, \mathbf{t}_{j-2}) - \mathsf{p}^*| \\
&\quad \times \|\mathsf{p}_{E_{j-1},\mathbf{t}_{j-2}} - \mathsf{p}^*\| \\
&\le \mathbf{Adv}^{\mathrm{ncpa}}_{E_{j-1}}(q) \sum_{\substack{\mathbf{t}_1 \in (\mathcal{M})_q : \\ C_{k_1,1}}} |\mathsf{p}_{E_1}(\mathbf{t}_0, \mathbf{t}_1) - \mathsf{p}^*| \ldots \sum_{\substack{\mathbf{t}_{j-2} \in (\mathcal{M})_q : \\ C_{k_{j-2},j-2}}} |\mathsf{p}_{E_{j-2}}(\mathbf{t}_{j-3}, \mathbf{t}_{j-2}) - \mathsf{p}^*| \\
&\vdots \\
&\le \prod_{2 \le i \le j-1} \mathbf{Adv}^{\mathrm{ncpa}}_{E_i}(q) \sum_{\substack{\mathbf{t}_1 \in (\mathcal{M})_q : \\ C_{k_1,1}}} |\mathsf{p}_{E_1}(\mathbf{t}_0, \mathbf{t}_1) - \mathsf{p}^*| \\
&\le \prod_{2 \le i \le j-1} \mathbf{Adv}^{\mathrm{ncpa}}_{E_i}(q) \times \|\mathsf{p}_{E_1,\mathbf{t}_0} - \mathsf{p}^*\| \\
&\le \prod_{1 \le i \le j-1} \mathbf{Adv}^{\mathrm{ncpa}}_{E_i}(q).
\end{aligned}$$

De même, on a :

$$S_2 = \sum_{\substack{\mathbf{t}_{n-1} \in (\mathcal{M})_q : \\ C_{k_n,n}}} |\mathsf{p}_{E_n}(\mathbf{t}_{n-1}, \mathbf{t}_n) - \mathsf{p}^*| \ldots \sum_{\substack{\mathbf{t}_j \in (\mathcal{M})_q : \\ C_{k_{j+1},j+1}}} |\mathsf{p}_{E_{j+1}}(\mathbf{t}_j, \mathbf{t}_{j+1}) - \mathsf{p}^*|$$

$$
\begin{aligned}
&\leq \sum_{\substack{t_{n-1}\in(\mathcal{M})_q:\\ C_{k_n,n}}} |\mathsf{p}_{E_n}(\mathbf{t}_{n-1},\mathbf{t}_n)-\mathsf{p}^*|\dots \sum_{\substack{\mathbf{t}_{j+1}\in(\mathcal{M})_q:\\ C_{k_{j+2},j+2}}} |\mathsf{p}_{E_{j+2}}(\mathbf{t}_{j+1},\mathbf{t}_{j+2})-\mathsf{p}^*| \\
&\qquad \times \|\mathsf{p}_{E_{j+1}^{-1},\mathbf{t}_{j+1}}-\mathsf{p}^*\| \\
&\leq \mathbf{Adv}^{\mathrm{ncpa}}_{E_{j+1}^{-1}}(q) \sum_{\substack{\mathbf{t}_{n-1}\in(\mathcal{M})_q:\\ C_{k_n,n}}} |\mathsf{p}_{E_n}(\mathbf{t}_{n-1},\mathbf{t}_n)-\mathsf{p}^*|\dots \sum_{\substack{\mathbf{t}_{j+1}\in(\mathcal{M})_q:\\ C_{k_{j+2},j+2}}} |\mathsf{p}_{E_{j+2}}(\mathbf{t}_{j+1},\mathbf{t}_{j+2})-\mathsf{p}^*| \\
&\vdots \\
&\leq \prod_{j+1\leq i\leq n} \mathbf{Adv}^{\mathrm{ncpa}}_{E_i^{-1}}(q),
\end{aligned}
$$

et il en suit l'inégalité voulue. □

1.5.5 Conclusion

D'après le lemme le lemme 4, on a

$$
\begin{aligned}
\mathsf{p}_{E_n\circ\cdots\circ E_1}(\mathbf{x},\mathbf{y}) &= \mathsf{p}^* + \sum_{(\mathbf{x}_1,\dots,\mathbf{x}_{n-1})\in((\mathcal{M})_q)^{n-1}} \left(\prod_{1\leq i\leq n} (\mathsf{p}_{E_i}(\mathbf{x}_{i-1},\mathbf{x}_i)-\mathsf{p}^*)\right) \\
&= \mathsf{p}^* + \sum_{\mathbf{k}\in\{0,1\}^n} \sum_{\substack{(\mathbf{x}_1,\dots,\mathbf{x}_{n-1})\in\\ A_{\mathbf{k}}(\mathbf{x}_0,\mathbf{x}_n)}} \left(\prod_{1\leq i\leq n} (\mathsf{p}_{E_i}(\mathbf{x}_{i-1},\mathbf{x}_i)-\mathsf{p}^*)\right) \\
&\geq \mathsf{p}^* + \sum_{\substack{\mathbf{k}\in\{0,1\}^n:\\ k_1+\dots+k_n\equiv 1 \bmod 2}} \sum_{\substack{(\mathbf{x}_1,\dots,\mathbf{x}_{n-1})\in\\ A_{\mathbf{k}}(\mathbf{x}_0,\mathbf{x}_n)}} \left(\prod_{1\leq i\leq n} (\mathsf{p}_{E_i}(\mathbf{x}_{i-1},\mathbf{x}_i)-\mathsf{p}^*)\right) \\
&\geq \mathsf{p}^*\left(1-2^{n-1}\max_{1\leq i\leq n}\left(\prod_{1\leq j\leq i-1}\mathbf{Adv}^{\mathrm{ncpa}}_{E_j}(q)\prod_{i+1\leq j\leq n}\mathbf{Adv}^{\mathrm{ncpa}}_{E_j^{-1}}(q)\right)\right),
\end{aligned}
$$

où, pour la dernière égalité, on a utilisé le lemme 5. En utilisant les égalités 1.2 et 1.2, on a

$$
\frac{\Pr[T_{\mathrm{re}}=\tau]}{\Pr[T_{\mathrm{id}}=\tau]} \geq 1-2^{n-1}\max_{1\leq i\leq n}\left(\prod_{1\leq j\leq i-1}\mathbf{Adv}^{\mathrm{ncpa}}_{E_j}(q)\prod_{i+1\leq j\leq n}\mathbf{Adv}^{\mathrm{ncpa}}_{E_j^{-1}}(q)\right),
$$

puis, en appliquant le lemme 1 avec $\Theta_{\mathrm{bad}}=\varnothing$ et $\varepsilon_2=0$, on obtient

$$
\mathbf{Adv}(\mathcal{D}) \leq 2^{n-1}\max_{1\leq i\leq n}\left(\prod_{1\leq j\leq i-1}\mathbf{Adv}^{\mathrm{ncpa}}_{E_j}(q)\prod_{i+1\leq j\leq n}\mathbf{Adv}^{\mathrm{ncpa}}_{E_j^{-1}}(q)\right)
$$

et le résultat suit en prenant le maximum sur les distingueurs cca effectuant q requêtes.

1.5.6 De l'exactitude de notre borne

Maurer, Pietrzak et Renner ont prouvé que le théorème 2W1S est optimal dans [MPR07] (voir l'annexe A de la version complète de [MPR07]). Nous allons généraliser cette preuve au corollaire 2 et prouver qu'il est optimal à une constante près.

Comme dans [MPR07], notons G la famille de toutes les permutations de $\mathcal{M}$ telles que 0 est dans un cycle de longueur 2 (c'est-à-dire que $\forall g \in G, g(g(0)) = 0$). En considérant G comme un algorithme de chiffrement par blocs (en ignorant les problèmes d'efficacité, ceci signifie simplement que l'on définit l'ensemble des clés par $\mathcal{K} = G$), on peut montrer que $\mathbf{Adv}_G^{\mathrm{ncpa}}(q) \leq \frac{2q}{|\mathcal{M}|}$ et $\mathbf{Adv}_G^{\mathrm{cca}}(2) \geq 1 - \frac{2}{|\mathcal{M}|}$. Définissons à présent l'algorithme de chiffrement par blocs F tel que :

— avec une probabilité ϵ, F est l'identité $\mathcal{I}$,

— avec une probabilité $1 - \epsilon$, F est choisi uniformément aléatoirement dans G.

Fixons également des réels $\delta, \delta', \delta'' > 0$. Alors

$$\mathbf{Adv}_F^{\mathrm{ncpa}}(q) = \varepsilon \mathbf{Adv}_{\mathcal{I}}^{\mathrm{ncpa}}(q) + (1-\varepsilon)\mathbf{Adv}_G^{\mathrm{ncpa}}(q) \leq \varepsilon + \frac{2q}{|\mathcal{M}|} \leq (1+\delta)\varepsilon, \qquad (1.7)$$

où, pour la dernière inégalité, on a supposé $|\mathcal{M}|$ suffisamment élevé.

Considérons maintenant l'algorithme de chiffrement par blocs F^n pour un entier naturel fixé $n \geq 2$. Considérons le distingueur adaptatif $\mathcal{D}$ qui effectue deux requêtes à son oracle de permutation P, $P(0)$ et ensuite $P(P(0))$, puis renvoie 1 *si et seulement si* $P(P(0)) = 0$. Si $\mathcal{D}$ interagit avec une permutation aléatoire, $\mathcal{D}$ renvoie 1 avec une probabilité égale à $2/|\mathcal{M}|$. En effet, avec probabilité $1/|\mathcal{M}|$, 0 est un point fixe de P, et avec probabilité $(|\mathcal{M}| - 1)/(\mathcal{M}|(|\mathcal{M}| - 1))$, on a $P(0) = y$ et $P(y) = 0$ pour une valeur $y \neq 0$. Dans le cas où $\mathcal{D}$ interagit avec F^n, il renvoie 1 en particulier lorsque $n-1$ des n instances de F sont l'identité, ce qui arrive avec probabilité $n(1-\varepsilon)\varepsilon^{n-1}$. Ainsi, pour tout $q \geq 2$, on a

$$\mathbf{Adv}_{F^n}^{\mathrm{cca}}(q) \geq n(1-\varepsilon)\varepsilon^{n-1} - \frac{2}{|\mathcal{M}|} \geq \frac{n}{(1+\delta')(1+\delta'')}\varepsilon^{n-1},$$

où, pour la dernière inégalité, on a supposé ε suffisamment petit et $|\mathcal{M}|$ suffisamment grand. En utilisant (1.7), on obtient finalement

$$\mathbf{Adv}_{F^n}^{\mathrm{cca}}(q) \geq \frac{n}{(1+\delta)^{n-1}(1+\delta')(1+\delta'')}(\mathbf{Adv}_F^{\mathrm{ncpa}})^{n-1}.$$

Puisqu'il est possible de rendre δ, δ', et δ'' aussi petits que l'on veut, ceci montre que, essentiellement, le meilleur majorant que l'on puisse espérer dans le corollaire 2 est $n\varepsilon^{n-1}$. Réduire l'écart entre la borne prouvée $2^{n-1}\varepsilon^{n-1}$ et $n\varepsilon^{n-1}$ reste un problème ouvert intéressant.

1.6 Illustration 2 : la construction EDM

1.6.1 Objectif

Dans la section précédente, nous avons utilisé le problème de l'amplification de sécurité pour la composition d'algorithmes de chiffrement par blocs pour illustrer

l'utilisation du lemme 1. Dans cette section, nous allons utiliser un problème différent pour expliciter l'utilisation de l'ensemble des mauvaises transcriptions : l'étude de la sécurité de la construction EDM.

Commençons par présenter ce nouveau problème. Les algorithmes de chiffrement par blocs sont omniprésents en cryptographie. Comme nous l'avons vu, le prérequis standard de sécurité pour un algorithme de chiffrement par blocs est qu'il soit indistinguable d'une permutation aléatoire [GGM86], ce qui signifie qu'aucun adversaire avec un accès en boîte noire à une permutation et avec des ressources limitées ne devrait être capable de distinguer avec un avantage non-négligeable s'il interagit avec l'algorithme de chiffrement par blocs équipé d'une clé choisie uniformément aléatoirement ou avec une permutation choisie uniformément aléatoirement.

La bijectivité peut sembler être un prérequis basique lorsque l'on utilise un algorithme de chiffrement par blocs pour chiffrer des données. Toutefois, cette intuition peut s'avérer fausse pour de nombreux modes d'opération. Par exemple, considérons le chiffrement en utilisant le mode compteur [BDJR97] appliqué à un algorithme de chiffrement par blocs d'espace de messages $\{0,1\}^n$: le message à chiffrer est découpé en blocs de message de n bits m_i qui sont chiffrés en $y_i = m_i \oplus E_k(c_i)$, où c_i est un compteur qui ne se répète pas. On sait que ce mode d'opération n'est sûr que tant que moins de $2^{n/2}$ blocs sont chiffrés à l'aide de la même clé. Dans ce cas, les chiffrés peuvent être distingués d'un message aléatoire par l'adversaire ; en effet, pour des y_i choisis uniformément aléatoirement, à cause du paradoxe des anniversaires, l'adversaire s'attend à trouver des collisions parmi les valeurs $y_i \oplus m_i$, alors que ceci n'arrive jamais pour la sortie d'un oracle de chiffrement réel puisque $y_i \oplus m_i = E_k(c_i)$, où les valeurs c_i sont toutes deux à deux distinctes. Ce problème disparaît si, au lieu d'utiliser un algorithme de chiffrement par blocs, on utilise une fonction pseudo-aléatoire. Tout comme un algorithme de chiffrement par blocs, une fonction pseudo-aléatoire F prend en entrée une clé $k \in \mathcal{K}$ et un texte clair $x \in \mathcal{X}$, et retourne un texte « chiffré » y dans un ensemble de sortie $\mathcal{Y}$, qui peut éventuellement être différent de l'espace $\mathcal{X}$ des messages clairs. Le prérequis de sécurité est maintenant qu'aucun adversaire ayant un accès en boîte noire à une fonction de $\mathcal{X}$ vers $\mathcal{Y}$ ne devrait pouvoir distinguer s'il interagit avec $F_k = F(k, \cdot)$ pour une clé k choisie aléatoirement ou avec une fonction uniformément aléatoire de $\mathcal{X}$ vers $\mathcal{Y}$ (au lieu d'une permutation aléatoire de $\mathcal{X}$ dans le cas d'un algorithme de chiffrement par blocs). On peut facilement constater qu'utiliser une fonction pseudo-aléatoire en mode compteur (c'est-à-dire chiffrer des blocs de message comme $y_i = m_i \oplus F_k(c_i)$) donne un mode de chiffrement qui préserve la sécurité de la fonction pseudo-aléatoire sous-jacente, dans le sens où l'avantage de tout adversaire contre le mode d'opération est majoré par son avantage contre la fonction pseudo-aléatoire elle-même, le mode en lui-même n'entraîne aucune dégradation de la sécurité (ce qui n'est pas le cas en utilisant un algorithme de chiffrement par blocs).

Un autre exemple d'application que l'on peut citer est celui des constructions MAC de Wegman-Carter [WC81], qui reposent sur une fonction pseudo-aléatoire F et une famille H de fonctions de hachage ε-AXU, une notion sur laquelle nous reviendrons dans le chapitre suivant, pour construire un code d'authentification de

message qui utilise un nonce défini par :

$$\mathsf{WC}[F, H](\nu, m) = F_k(\nu) \oplus H_{k'}(m),$$

où ν est le nonce (une valeur qui ne devrait jamais être répétée) et m le message à authentifier. La construction de Wegman-Carter jouit d'une excellente borne de sécurité lorsqu'elle est utilisée avec une bonne fonction pseudo-aléatoire et une bonne famille de fonctions de hachage : pour des codes authentificateurs de n bits, la probabilité qu'a un adversaire de forger peut être proche de $q_f/2^n$, où q_f est le nombre de tentatives de forge de l'adversaire, ce à quoi il faut ajouter un terme reliée à la sécurité de F. Cependant, si F est remplacée par un algorithme de chiffrement par blocs, la sécurité prouvée retombe au niveau de la borne des anniversaires [Sho96, Ber05].

Ces deux exemples montrent que la bijectivité peut nuire à la sécurité pour de nombreuses constructions fondées sur des algorithmes de chiffrement par blocs. Malheureusement, les cryptologues se sont concentrés sur la conception de bons algorithmes de chiffrement par blocs, et les fonctions pseudo-aléatoires sûres et efficaces ne sont pas courantes. Ainsi, une question naturelle se pose : est-il possible de convertir un algorithme de chiffrement par blocs E en une fonction pseudo-aléatoire $F[E]$ aussi efficacement que possible et en préservant la sécurité de E (c'est-à-dire que l'avantage d'un adversaire $\mathcal{D}$ pour distinguer $F[E]$ d'une fonction aléatoire devrait être proche de l'avantage d'un adversaire $\mathcal{D}'$ lié à $\mathcal{D}$ et disposant de ressources similaires pour distinguer E d'une permutation aléatoire, sans dégradation supplémentaire de la sécurité) ? Remarquons que tout algorithme de chiffrement par blocs E est une fonction pseudo-aléatoire sûre, toutefois sa sécurité est limitée à ce que l'on appelle la borne des anniversaires, c'est-à-dire jusqu'à environ $2^{n/2}$ requêtes de l'adversaire, même si E est sûre, en tant que permutation pseudo-aléatoire, face à un nombre bien plus élevé de requêtes. En effet, pour un tel nombre de requêtes, l'adversaire s'attend à observer, avec une probabilité élevée, des collisions entre les sorties d'une fonction aléatoire, alors que cela n'arrive jamais dans le cas d'une permutation. Ce résultat est souvent appelé le lemme d'échange PRP/PRF (pour *pseudorandom function* et *pseudorandom permutation*) [BR06b]. Ainsi, une méthode qui convertit un algorithme de chiffrement par blocs en fonction pseudo-aléatoire doit, au minimum, dépasser la borne des anniversaires pour présenter un intérêt.

Le problème réciproque, qui est de construire une fonction pseudo-aléatoire en algorithme de chiffrement par blocs a été résolu il y a environ 30 ans dans un article célèbre par Luby et Rackoff [LR88] en utilisant un schéma de Feistel à 3 tours. Si l'on souhaite un algorithme de chiffrement par blocs sûr face à des adversaires effectuant des requêtes bidirectionnelles, 4 tours sont nécessaires [LR88, Pat90]). Pour cette raison, le problème de la conversion de fonctions pseudo-aléatoires en algorithme de chiffrement par blocs est parfois appelé « Luby-Rackoff à l'envers » [BKR98].

Un nombre significatif de constructions ont été suggérées pour résoudre de problème. Une des plus simple est peut-être la troncature : on ignore m bits de la sortie de l'algorithme de chiffrement par blocs et on n'utilise que, par exemple, les $n - m$ bits les plus significatifs de la sortie de E_k pour la sortie de $F[E]_k$. Cette construction a été analysée par Hall *et al.* [HWKS98], qui ont montré qu'elle est sûre tant que le

nombre de requêtes de l'adversaire est petit devant $\min\{2^{(n+m)/2}, 2^{2(n-m)/3}\}$ (cette borne a ensuite été améliorée par Bellare et Impagliazzo [BI99]). Dans le même article [HWKS98], Hall *et al.* ont également étudié une construction inefficace mais qui préserve la sécurité fondée sur le classement des sorties $E_k(1\|x), \ldots, E_k(d\|x)$. Notons que ces constructions ne préservent pas l'ensemble image (ni l'ensemble de définition, dans le cas de la dernière construction) de la permutation originale.

Une autre option suggérée par Bellare, Krovetz, et Rogaway [BKR98] est de modifier la clé de l'algorithme de chiffrement par blocs selon les données. Dans le cas simple où l'espace de clés $\mathcal{K}$ de l'algorithme de chiffrement par blocs est identique à son espace de messages $\mathcal{X}$, cette construction est définie par $F(k, x) = E(E(k, x), x)$. Elle offre également de fortes garanties de sécurité, au-delà de la borne des anniversaires, mais uniquement en modélisant E comme un algorithme de chiffrement par blocs choisi uniformément aléatoirement. Dans le modèle standard, la borne de sécurité retombe au niveau de la borne des anniversaires.

Une autre méthode simple, que nous appelons la construction XOR, consiste simplement à ajouter les sorties de $r \geq 2$ chiffrements indépendants de l'entrée. Plus précisément, en supposant que l'espace de messages de E est $\{0,1\}^n$,

$$F_{(k_1,\ldots,k_r)}(x) = E_{k_1}(x) \oplus E_{k_2}(x) \oplus \cdots \oplus E_{k_r}(x),$$

où $k_1, \ldots, k_r$ sont des clés indépendantes. Cette construction a d'abord été analysée par Lucks [Luc00] qui a prouvé qu'elle était sûre tant que le nombre de requêtes de l'adversaire reste petit devant $2^{rn/(r+1)}$. Simultanément, Bellare et Impagliazzo [BI99] ont indépendamment démontré que, pour $r = 2$, l'avantage d'un adversaire est majoré par $O(n)(q/2^n)^{3/2}$; en d'autres termes, la sécurité est garantie tant que le nombre de requêtes de l'adversaire reste petit devant $2^n/n^{2/3}$. Patarin [Pat08a, Pat13] a prouvé, en utilisant deux variantes différentes de la technique des coefficients H, que la construction pour $r = 2$ est déjà « optimale », c'est à dire qu'elle est sûre tant que le nombre de requêtes de l'adversaire est petit devant 2^n. Le cas $r > 2$ a été peu étudié depuis [Luc00]. En effet, l'utilisation d'un nombre plus élevé de permutations ne peut qu'augmenter la sécurité qui est déjà optimale pour $r = 2$. Cependant, si la sécurité dans ce cas est optimale en terme de requêtes de l'adversaire, il est possible qu'utiliser un nombre plus élevé de permutations entraîne un gain géométrique en terme d'ordre de grandeur de l'avantage maximal que peut avoir un adversaire pour distinguer cette construction d'une fonction aléatoire. Rodolphe Lampe, Jacques Patarin et moi avons fait un premier pas dans ce sens lorsque l'attaquant est limité à un « petit » nombre de requêtes [CLP14]. La construction XOR peut être légèrement modifiée pour n'utiliser qu'une seule clé avec une perte négligeable de sécurité en définissant ce que Lucks appelle la construction TWIN [Luc00], définie par

$$F_k(x) = E_k(0\|x) \oplus E_k(1\|x) \oplus E_k(r-1\|x).$$

Cette construction réduit cependant légèrement l'ensemble de définition de la fonction pseudo-aléatoire de $\lceil \log_2(r) \rceil$ bits.

Une autre idée naturelle pour convertir un algorithme de chiffrement par blocs E en fonction pseudo-aléatoire est de définir $F_k(x) = E_k(x) \oplus x$. Quand F est vue comme une fonction de compression de $2n$ bits vers n bits, elle est appelée la

construction de Davies-Meyer (DM). Nous utiliserons également cette terminologie ici, bien que nous nous concentrions sur les fonctions pseudo-aléatoires et non les fonctions de hachage. Si la construction supprime bel et bien la bijectivité de E_k, il est facile de voir que cette construction n'est pas plus sûre que E. En effet, l'adversaire peut simplement calculer $F_k(x) \oplus x = E_k(x)$ et ainsi appliquer l'attaque standard de recherche de collision. L'idée la plus simple pour pallier ce défaut, et que nous allons considérer ici, est de chiffrer à nouveau la sortie de la construction DM (sous une clé indépendante) : cette nouvelle construction sera appelée la construction de Davies-Meyer chiffrée et notée EDM (pour *Encrypted Davies-Meyer*). Plus précisément, soient E et F deux algorithmes de chiffrement par blocs d'espaces de clés respectifs $\mathcal{K}_E$ et $\mathcal{K}_F$ et d'espace de messages $\{0,1\}^n$. La construction EDM est définie, pour tout $(k, k', x) \in \mathcal{K}_E \times \mathcal{K}_F \times \{0,1\}^n$, par

$$\mathsf{EDM}[E,F]_{k,k'}(x) = F_{k'}(E_k(x) \oplus x).$$

Nous noterons $\mathsf{EDM}[P, P']$ la construction EDM utilisée avec deux permutations uniformément aléatoires et indépendantes P et P'.

1.6.2 Notations et description des transcriptions

Fixons pour le reste de cette section un entier naturel n et deux algorithmes de chiffrement par blocs E et F d'espaces de clés respectifs $\mathcal{K}_E$ et $\mathcal{K}_F$ et d'ensemble de messages $\{0,1\}^n$. Soit $N = 2^n$. On notera $\mathsf{Func}(n)$ l'ensemble des fonctions de $\{0,1\}^n$ vers $\{0,1\}^n$ et $\mathsf{Perm}(n)$ l'ensemble des permutations de $\{0,1\}^n$.

Nous avons déjà présenté la notion d'indistinguabilité appliquée aux algorithmes de chiffrement par blocs. Ici, nous n'étudions pas la capacité d'un adversaire à distinguer une construction d'une permutation uniformément aléatoire, mais plutôt d'une fonction uniformément aléatoire. Nous allons donc devoir modifier légèrement notre modélisation de la situation : les adversaires ne seront autorisés qu'à effectuer des requêtes directes à leur oracle. Les classes d'attaques ncca et cca vont donc disparaître et nous allons considérer la résistance de la construction EDM face aux distingueurs cpa.

Formellement, un adversaire $\mathcal{D}$ qui essaye de distinguer $\mathsf{EDM}[E,F]$ utilisé avec des clés choisies uniformément aléatoirement, d'une fonction uniformément aléatoire $R \leftarrow_{\$} \mathsf{Func}(n)$ est vu comme un algorithme déterministe ayant une puissance de calcul non-bornée et disposant d'un oracle noté F de façon générique. Cet oracle simule soit R dans le monde idéal, soit $\mathsf{EDM}[E,F]$ dans le monde réel. Le distingueur $\mathcal{D}$ choisit $x \in \{0,1\}^n$, puis F lui donne sa réponse $F(x) \in \{0,1\}^n$. Après exactement q requêtes, $\mathcal{D}$ répond $\mathcal{D}^F \in \{0,1\}$. Afin de mesurer la nature pseudo-aléatoire de la construction EDM, il faut évaluer l'avantage $\mathbf{Adv}_{\mathsf{EDM}[E,F]}(\mathcal{D})$ de l'adversaire $\mathcal{D}$ qui est définie, comme précédemment, par

$$\mathbf{Adv}_{\mathsf{EDM}[E,F]}(\mathcal{D}) = |Pr[\mathcal{D}^{\mathsf{EDM}[E,F]} = 1] - Pr[\mathcal{D}^R = 1]|.$$

Nous écrirons, comme d'habitude, $\mathbf{Adv}^{\mathrm{cpa}}_{\mathsf{EDM}[E,F]}$ pour l'avantage maximal qu'un adversaire cpa peut obtenir lorsqu'il souhaite distinguer la construction $\mathsf{EDM}[E,F]$ d'une fonction aléatoire.

Comme dans la section précédente, la transcription τ de l'interaction entre le distingueur et son oracle F consiste simplement en une liste de paires (x, y) telles que, si $(x, y) \in \tau$, alors le distingueur a effectué la requête directe $x \in \{0,1\}^n$ et reçu en guise de réponse $F(x) = y$. Une transcription est dite atteignable s'il existe une fonction $R \in \mathsf{Func}(n)$ telle que l'interaction de $\mathcal{D}$ avec R donne la transcription τ. On note Θ l'ensemble des transcriptions atteignables. On note également T_{re}, respectivement T_{id}, la loi de probabilité de la transcription τ induite par le monde réel, respectivement le monde idéal.

1.6.3 Résultat

Avant d'entamer l'étude proprement dite de la construction EDM, la première étape consiste à remplacer les algorithmes de chiffrement par blocs équipés de clés choisies uniformément aléatoirement E et F par des permutations uniformément aléatoires et indépendantes P, P'. En effet, intuitivement, si les permutations pseudo-aléatoires $E_k, F_{k'}$ sont indistinguables de permutations uniformément aléatoires et indépendantes, alors la sécurité de la construction EDM dans ce cas de figure sera très proche de celle de la construction EDM utilisée avec des permutations aléatoires. Cependant, l'étude mathématique de la construction deviendra beaucoup plus simple à effectuer. Cette intuition est formalisée dans le lemme suivant.

Lemme 6. *Pour tout entier naturel q, on a*

$$\mathbf{Adv}^{\mathrm{cpa}}_{\mathsf{EDM}[E,F]}(q) \leq \mathbf{Adv}^{\mathrm{cpa}}_{E}(q) + \mathbf{Adv}^{\mathrm{cpa}}_{F}(q) + \mathbf{Adv}^{\mathrm{cpa}}_{\mathsf{EDM}[P,P']}(q).$$

Démonstration. Ce résultat s'obtient par un argument hybride classique. Soit $\mathcal{D}$ un attaquant cpa essayant de distinguer $\mathsf{EDM}[E, F]$ d'une fonction uniformément aléatoire. Définissons l'attaquant $\mathcal{D}_E$, respectivement $\mathcal{D}_F$, qui cherche à distinguer l'algorithme de chiffrement par blocs E, respectivement F, équipé d'une clé uniformément aléatoire, d'une permutation aléatoire. Le distingueur $\mathcal{D}_E$ exécute $\mathcal{D}$ et répond à ses requêtes de la façon suivante : il calcule la sortie de la construction EDM appliquée à la requête de $\mathcal{D}$, en utilisant, pour la première permutation son propre oracle de permutation, puis, pour la seconde permutation, il évalue l'algorithme de chiffrement par blocs F équipé d'une clé qu'il aura tirée uniformément aléatoirement. Il renverra ensuite la même sortie que l'attaquant $\mathcal{D}$. Son avantage pour distinguer E d'une permutation aléatoire vaut donc exactement :

$$\mathbf{Adv}_{E}(\mathcal{D}_E) = \left|\Pr\left[\mathcal{D}^{\mathsf{EDM}[E,F]} = 1\right] - \Pr\left[\mathcal{D}^{\mathsf{EDM}[P,F]} = 1\right]\right|.$$

De façon similaire, le distingueur $\mathcal{D}_F$ exécute $\mathcal{D}$ et répond à ses requêtes de la façon suivante : il calcule la sortie de la construction EDM appliquée à la requête de $\mathcal{D}$, en utilisant, pour la première permutation, une permutation uniformément aléatoire P, puis, pour la seconde permutation, il utilise son propre oracle de permutation. Il renverra ensuite la même sortie que l'attaquant $\mathcal{D}$. Son avantage pour distinguer F d'une permutation aléatoire vaut donc exactement :

$$\mathbf{Adv}_{F}(\mathcal{D}_F) = \left|\Pr\left[\mathcal{D}^{\mathsf{EDM}[P,F]} = 1\right] - \Pr\left[\mathcal{D}^{\mathsf{EDM}[P,P']} = 1\right]\right|.$$

On remarquera que $\mathcal{D}_E$ et $\mathcal{D}_F$ sont adaptatifs et effectuent au plus q requêtes directes à leur propre oracle. Ainsi, par définition de l'avantage et en utilisant l'inégalité triangulaire, on a :

$$\begin{aligned}
\mathbf{Adv}_{\mathsf{EDM}[E,F]}(\mathcal{D}) =& |\Pr[\mathcal{D}^{\mathsf{EDM}[E,F]} = 1] - \Pr[\mathcal{D}^{F} = 1]| \\
=& |\Pr[\mathcal{D}^{\mathsf{EDM}[E,F]} = 1] - \Pr[\mathcal{D}^{\mathsf{EDM}[P,F]} = 1] + \Pr[\mathcal{D}^{\mathsf{EDM}[P,F]} = 1] \\
& \quad - \Pr[\mathcal{D}^{\mathsf{EDM}[P,P']} = 1] + \Pr[\mathcal{D}^{\mathsf{EDM}[P,P']} = 1] - \Pr[\mathcal{D}^{F} = 1]| \\
\leq& |\Pr[\mathcal{D}^{\mathsf{EDM}[E,F]} = 1] - \Pr[\mathcal{D}^{\mathsf{EDM}[P,F]} = 1]| \\
& \quad + |\Pr[\mathcal{D}^{\mathsf{EDM}[P,F]} = 1] - \Pr[\mathcal{D}^{\mathsf{EDM}[P,P']} = 1]| \\
& \quad + |\Pr[\mathcal{D}^{\mathsf{EDM}[P,P']} = 1] - \Pr[\mathcal{D}^{F} = 1]| \\
\leq& \mathbf{Adv}_{E}(\mathcal{D}_E) + \mathbf{Adv}_{F}(\mathcal{D}_F) + \mathbf{Adv}_{\mathsf{EDM}[P,P']}(\mathcal{D}),
\end{aligned}$$

d'où le résultat annoncé. □

Dans le reste de cette section, nous allons donc nous restreindre à l'étude de la construction $\mathsf{EDM}[P, P']$ où P et P' sont deux permutations aléatoires indépendantes de $\{0,1\}^n$, et nous allons prouver que cette construction est sûre tant que l'adversaire effectue un nombre de requêtes petit devant $2^{2n/3}$. Plus précisément, on a le théorème suivant.

Théorème 2 ([CS16a])**.** *Soit $\mathcal{D}$ un adversaire ayant accès à une fonction de $\{0,1\}^n$ vers $\{0,1\}^n$, effectuant au plus q requêtes, et renvoyant un bit unique. Alors son avantage pour distinguer la construction* EDM *d'une fonction uniformément aléatoire, défini par*

$$\left|\Pr\left[(P,P') \leftarrow_{\$} \mathsf{Perm}(n)^2 : \mathcal{D}^{\mathsf{EDM}[P,P']} = 1\right] - \Pr\left[R \leftarrow_{\$} \mathsf{Func}(n) : \mathcal{D}^{R} = 1\right]\right|,$$

est inférieur à $5q^{3/2}/2^n$.

La preuve de ce résultat repose sur la technique des coefficients H. Dans la sous-section suivante, nous allons définir l'ensemble des mauvaises transcriptions, puis majorer la probabilité qu'ils apparaissent dans le monde idéal. Nous étudierons ensuite les bonnes transcriptions. Fixons, pour le reste de la section, un adversaire $\mathcal{D}$ effectuant exactement q requêtes, pour un entier naturel $q \in \{1, \ldots, 2^n\}$.

1.6.4 Description des mauvaises transcriptions

Commençons par donner une définition des mauvaises transcriptions. Soit τ une transcription atteignable. On notera $\tau = ((x_1, y_1), \ldots, (x_q, y_q))$. On dit qu'une requête $(x_i, y_i) \in \tau$ est en collision si $y_i = y_j$ pour un indice $j \neq i$ et on dira dans le cas contraire qu'elle n'est pas en collision.

Définition 2. *On dit qu'une transcription atteignable τ est mauvaise si le nombre de requêtes qui sont en collision est supérieur à $\sqrt{q}$. Dans le cas contraire, on dira que τ est une bonne transcription. On notera Θ_{good}, respectivement Θ_{bad} l'ensemble des bonnes, respectivement des mauvaises transcriptions.*

Nous allons à présent majorer la probabilité d'obtenir une mauvaise transcription dans le monde idéal.

Lemme 7. *Pour tout entier naturel q, on a*

$$\Pr\left[T_{\mathrm{id}} \in \Theta_{\mathrm{bad}}\right] \leq \frac{q^{3/2}}{2^n}. \tag{1.8}$$

Démonstration. On rappelle que, dans le monde idéal, $\mathcal{D}$ interagit avec une fonction choisie uniformément aléatoirement et indépendamment de la transcription. Comme le distingueur n'effectue jamais deux fois la même requête, les valeurs y_i pour $i = 1, \ldots, q$ sont uniformément aléatoires et indépendantes. Notons C le nombre de requêtes en collision. Alors on a

$$\begin{aligned}\mathbb{E}\left[C\right] &\leq \sum_{i=1}^{q} \sum_{\substack{j=1 \\ j\neq i}}^{q} \Pr\left[y_i = y_j\right] \\ &\leq \frac{q^2}{2^n}.\end{aligned}$$

On a ainsi, d'après l'inégalité de Markov,

$$\begin{aligned}\Pr\left[T_{\mathrm{id}} \in \Theta_{\mathrm{bad}}\right] &= \Pr\left[C \geq \sqrt{q}\right] \\ &\leq \frac{\mathbb{E}\left[C\right]}{\sqrt{q}} \\ &\leq \frac{q^{3/2}}{2^n}.\end{aligned}$$

□

1.6.5 Étude des bonnes transcriptions

Fixons à présent une bonne transcription τ. Nous devons à présent minorer la probabilité d'obtenir τ dans le monde réel. On a le résultat suivant.

Lemme 8. *Soit q un entier naturel. Pour toute bonne transcription τ, on a*

$$\frac{\Pr\left[T_{\mathrm{re}} = \tau\right]}{\Pr\left[T_{\mathrm{id}} = \tau\right]} \geq 1 - \frac{4q^{3/2}}{2^n}. \tag{1.9}$$

Démonstration. Remarquons tout d'abord que, si $q^{3/2} > 2^n/4$, le résultat est trivial. Nous allons donc supposer que $q^{3/2} \leq 2^n/4$. En particulier, $q \leq 2^n/2$. Commençons par réordonner la transcription comme suit. Notons r le nombre de valeurs différentes dans le vecteur $(y_1, \ldots, y_q)$. Réécrivons alors la transcription de sorte que les requêtes partageant la même seconde coordonnée. La transcription devient alors

$$\begin{aligned}\tau = \Big(&(x_{1,1}, y_1), \ldots, (x_{1,q_1}, y_1), \\ &(x_{2,1}, y_2), \ldots, (x_{2,q_2}, y_2), \\ &\ldots, \\ &(x_{r,1}, y_r), \ldots, (x_{r,q_r}, y_r)\Big),\end{aligned}$$

où les valeurs $y_1, \ldots, y_r$ sont deux à deux distinctes et $\sum_{i=1}^{r} q_i = q$.

Dans le but de minorer la probabilité d'obtenir τ dans le monde réel, nous devons minorer le nombre de paires de permutations (P, P') telles que

$$\forall i \in \{1, \ldots, r\},\ \forall j \in \{1, \ldots, q_i\},\ P'(P(x_{i,j}) \oplus x_{i,j}) = y_i.$$

Dans ce but, nous allons considérer toutes les valeurs « internes » possibles $z_i = (P')^{-1}(y_i)$. On dit qu'un uplet $\mathbf{z} = (z_1, \ldots, z_r)$ de valeurs deux à deux distinctes est *bon* si toutes les valeurs $z_i \oplus x_{i,j}$ pour $i \in \{1, \ldots, r\}$ et $j \in \{1, \ldots, q_i\}$ sont deux à deux distinctes. Étant donné un bon uplet $\mathbf{z}$, la probabilité que

$$\begin{cases} \forall i \in \{1, \ldots, r\},\ \forall j \in \{1, \ldots, q_i\},\ P(x_{i,j}) = z_i \oplus x_{i,j}, \\ \forall i \in \{1, \ldots, r\},\ P'(z_i) = y_i \end{cases}$$

vaut exactement

$$\frac{1}{(2^n)_q (2^n)_r}. \tag{1.10}$$

En effet, il s'agit simplement de la probabilité que P vérifie $q_1 + \ldots + q_r = q$ équations et que P' vérifie r équations.

Nous pouvons minorer le nombre de bons uplets $\mathbf{z}$ comme suit :

— il y a au moins 2^n possibilités pour z_1 ;

— une fois z_1 fixé, il reste au moins $2^n - 1 - q_1 q_2$ possibilités pour z_2, puisque z_2 doit être différent de z_1 et de $z_1 \oplus x_{1,j} \oplus x_{2,j'}$ pour tout $j \in \{1, \ldots, q_1\}$ et tout $j' \in \{1, \ldots, q_2\}$;

— une fois z_1 et z_2 fixés, il reste au moins $2^n - 2 - (q_1 + q_2) q_3$ possibilités pour z_3, puisque z_3 doit être différent de z_1, z_2, $z_1 \oplus x_{1,j} \oplus x_{3,j'}$ pour tout $j \in \{1, \ldots, q_1\}$ et tout $j' \in \{1, \ldots, q_3\}$, mais aussi de $z_2 \oplus x_{2,j} \oplus x_{3,j'}$ pour tout $j \in \{1, \ldots, q_2\}$ et tout $j' \in \{1, \ldots, q_3\}$;

— etc.

Ainsi, le nombre de bons uplets $\mathbf{z}$ vaut au moins

$$\prod_{i=0}^{r-1} \left(2^n - i - q_{i+1} \sum_{j=1}^{i} q_j \right).$$

Par conséquent, en ajoutant la probabilité (1.10) pour toutes les valeurs possibles pour $\mathbf{z}$, la probabilité d'obtenir la transcription τ dans le monde réel vérifie

$$\Pr\left[T_{\mathrm{re}} = \tau\right] \geq \frac{\prod_{i=0}^{r-1} \left(2^n - i - q_{i+1} \sum_{j=1}^{i} q_j \right)}{(2^n)_q (2^n)_r}.$$

Puisque la probabilité d'obtenir τ dans le monde idéal vaut simplement $1/(2^n)^q$, le rapport entre les deux probabilités vaut au moins

$$\begin{aligned} \rho \stackrel{\mathrm{def}}{=} \frac{\Pr\left[T_{\mathrm{re}} = \tau\right]}{\Pr\left[T_{\mathrm{id}} = \tau\right]} &\geq \frac{(2^n)^q \prod_{i=0}^{r-1} \left(2^n - i - q_{i+1} \sum_{j=1}^{i} q_j \right)}{(2^n)_q (2^n)_r} \\ &= \prod_{i=0}^{q-1} \left(1 + \frac{i}{2^n - i} \right) \prod_{i=0}^{r-1} \left(1 - \frac{q_{i+1} \sum_{j=1}^{i} q_j}{2^n - i} \right). \end{aligned}$$

Dans le but de minorer plus précisément le rapport ρ, nous devons différencier les requêtes qui sont en collision de celle qui ne le sont pas. Quittes à réordonner la transcription, nous pouvons supposer que les requêtes qui ne sont pas en collision apparaissent au début de la transcription, et notons $s \in \{0, \ldots, r\}$ l'unique entier tel que $q_i = 1$ pour $i \in \{1, \ldots, s\}$, et $q_i > 1$ pour $i \in \{s+1, \ldots, r\}$. Notons que, puisque la transcription est bonne,

$$\sum_{i=s+1}^{r} q_i \leq \sqrt{q}. \tag{1.11}$$

Alors,

$$\begin{aligned}
\rho &\geq \prod_{i=0}^{q-1} \left(1 + \frac{i}{2^n - i}\right) \prod_{i=0}^{s-1} \left(1 - \frac{q_{i+1} \sum_{j=1}^{i} q_j}{2^n - i}\right) \prod_{i=s}^{r-1} \left(1 - \frac{q_{i+1} \sum_{j=1}^{i} q_j}{2^n - i}\right) \\
&= \prod_{i=0}^{q-1} \left(1 + \frac{i}{2^n - i}\right) \prod_{i=0}^{s-1} \left(1 - \frac{i}{2^n - i}\right) \prod_{i=s}^{r-1} \left(1 - \frac{q_{i+1} \sum_{j=1}^{i} q_j}{2^n - i}\right) \\
&\geq \prod_{i=0}^{q-1} \left(1 - \frac{i^2}{(2^n - i)^2}\right) \prod_{i=s}^{r-1} \left(1 - \frac{q_{i+1} q}{2^n - i}\right) \\
&\geq \prod_{i=0}^{q-1} \left(1 - \frac{i^2}{(2^n - q)^2}\right) \prod_{i=s}^{r-1} \left(1 - \frac{q_{i+1} q}{2^n - q}\right) \\
&\geq \left(1 - \frac{q^3}{3(2^n - q)^2}\right) \left(1 - \frac{q \sum_{i=s+1}^{r} q_i}{2^n - q}\right) \\
&\geq \left(1 - \frac{4q^3}{3 \cdot 2^{2n}}\right) \left(1 - \frac{2q^{3/2}}{2^n}\right),
\end{aligned}$$

où, pour la dernière inégalité, on a utilisé le fait que $q \leq 2^n/2$ et (1.11). Puisque $q^{3/2} \leq 2^n/4$, on a en particulier que $q^3/2^{2n} \leq q^{3/2}/2^n$. Ainsi, on obtient

$$\frac{\Pr\left[T_{\mathrm{re}} = \tau\right]}{\Pr\left[T_{\mathrm{id}} = \tau\right]} \geq 1 - \frac{4q^{3/2}}{2^n}. \qquad \square$$

1.6.6 Conclusion

Nous pouvons à présent achever la preuve du théorème 2. Pour ce faire, il suffit de combiner les lemmes 7 et 8 avec le lemme 1. On obtient ainsi, comme annoncé, le fait que l'avantage du distingueur $\mathcal{D}$ est inférieur à $5q^{3/2}/2^n$.

Nous conjecturons que ce théorème n'est en fait pas optimal, et que la sécurité offerte par notre construction est similaire à celle offerte par la construction XOR, c'est-à-dire que l'avantage du distingueur $\mathcal{D}$ est inférieur, à une constante multiplicative près, à $q/2^n$.

1.7 Illustration 3 : l'utilisation de la construction EDM avec une permutation

1.7.1 Résultat

Dans la section précédente, nous avons prouvé que la construction $\mathsf{EDM}[P, P']$, lorsque P et P' sont deux permutations de l'ensemble des chaînes de caractères de n bits, est indistinguable d'une fonction uniformément aléatoire tant que le nombre de requêtes de l'adversaire est petit devant $2^{2n/3}$. Ainsi, si l'on veut utiliser cette construction en initialisant les permutations grâce à un algorithme de chiffrement par blocs, il est nécessaire d'utiliser deux clés choisies uniformément aléatoirement avec une distribution uniforme. Mais on peut se demander ce qui se passe si l'on utilise la même clé deux fois dans notre construction. Nous avons également étudié ce cas de figure et prouvé que, lorsque la même permutation est utilisée deux fois, la construction EDM offre un niveau de sécurité similaire à celui de la construction EDM utilisée avec deux permutations indépendantes.

Plus précisément, dans cette section, nous allons considérer la construction de Davies-Meyer chiffrée (à une permutation)

$$\mathsf{EDM}[P](x) = P(P(x) \oplus x),$$

où P est une permutation aléatoire de $\{0,1\}^n$. Alors nous avons le théorème suivant.

Théorème 3. *Soient $n \geq 6$ et $q \in \{36, \ldots, \left\lfloor \frac{2^n}{6} \right\rfloor\}$. Soit $\mathcal{D}$ un adversaire ayant un accès en boîte noire à une fonction de $\{0,1\}^n$ vers $\{0,1\}^n$, effectuant au plus q requêtes, et renvoyant un unique bit. Alors son avantage pour distinguer la construction EDM d'une fonction uniformément aléatoire, défini comme*

$$\mathbf{Adv}(\mathcal{D}) \stackrel{\text{def}}{=} \left| \Pr\left[P \leftarrow_{\$} \mathsf{Perm}(n) : \mathcal{D}^{\mathsf{EDM}[P]} = 1\right] - \Pr\left[R \leftarrow_{\$} \mathsf{Func}(n) : \mathcal{D}^{R} = 1\right] \right|,$$

vérifie :

$$\mathbf{Adv}(\mathcal{D}) \leq \frac{52q}{2^{2n/3}} + \frac{8\sqrt{3nq}}{2^{n/3}}.$$

Comme dans la section précédente, on obtient directement le corollaire suivant lorsque la construction est utilisée avec un algorithme de chiffrement par blocs équipé d'une clé choisie uniformément aléatoirement.

Corollaire 3. *Soit E un algorithme de chiffrement par blocs d'espace de clés $\mathcal{K}$ et d'espace de messages $\{0,1\}^n$. Soient $n \geq 6$ et $q \in \{36, \ldots, \left\lfloor \frac{2^n}{6} \right\rfloor\}$ deux entiers naturels. Alors pour tout adversaire $\mathcal{D}$ effectuant au plus q requêtes et dont le but est de distinguer $\mathsf{EDM}[E]$ d'une fonction aléatoire, il existe un adversaire $\mathcal{D}'$ effectuant au plus $2q$ requêtes et dont le but est de distinguer E d'une permutation aléatoire tel que :*

$$\mathbf{Adv}_{\mathsf{EDM}[E]}(\mathcal{D}) \leq \mathbf{Adv}_{E}(\mathcal{D}') + \frac{52q}{2^{2n/3}} + \frac{8\sqrt{3nq}}{2^{n/3}}.$$

Le reste de la section sera consacré à la preuve du théorème 3. Le but de cette section est d'expliciter une stratégie de preuve plus complexe reposant sur la technique des coefficients H. Les preuves de quatre lemmes calculatoires seront donc omises et pourront être trouvées dans l'annexe B de ce manuscrit. Avant de débuter la démonstration, constatons que le résultat est évident si $q > 2^{2n/3}$. Nous allons donc fixer, pour le reste de cette section, deux entiers naturels n et q tels que $n \geq 6$, $36 \leq q \leq 2^n/6$ et $q \leq 2^{2n/3}$. Ces conditions entraînent en particulier que

$$\frac{q^3}{2^n} \leq q^{3/2} \leq 2^n. \tag{1.12}$$

1.7.2 Description des mauvaises transcriptions

Fixons un distingueur $\mathcal{D}$ effectuant q requêtes et commençons par décrire les mauvaises transcriptions.

Définition 3. *On dit qu'une transcription atteignable τ est mauvaise si l'une des conditions suivantes est vérifiée :*

(i) le nombre de requêtes en collision est supérieur à $\sqrt{q}$;

(ii) le nombre $\alpha(\tau)$ de triplets $(y, x, x') \in Y^* \times X \times X$ tels que $y = x \oplus x'$ est supérieur à $q^3/2^n + q\sqrt{3nq}$.

Dans le cas contraire, on dit que τ est une bonne transcription. On notera Θ_{bad}, respectivement Θ_{good}, l'ensemble des mauvaises, respectivement des bonnes transcriptions.

On majore ensuite la probabilité d'obtenir une mauvaise transcription dans le monde idéal. On a le lemme suivant, dont la preuve peut être trouvée dans l'annexe B.2.

Lemme 9. *Supposons $q \geq 2$. Alors*

$$\Pr\left[T_{\text{id}} \in \Theta_{\text{bad}}\right] \leq \frac{2q^{3/2}}{2^n}.$$

1.7.3 Étude des bonnes transcriptions

À présent, fixons une bonne transcription $\tau = ((x_1, y_1), \ldots, (x_q, y_q))$. Soient $X = \{x_1, \ldots, x_q\}$ et $Y = \{y_1, \ldots, y_q\}$. Nous avons maintenant besoin de minorer la probabilité d'observer τ dans le monde réel. Comme précédemment, cette probabilité correspond exactement à la probabilité que l'oracle du monde réel soit « compatible » avec la transcription (voir par exemple [CS14]), c'est-à-dire

$$\Pr\left[T_{\text{id}} = \tau\right] = \Pr\left[P \leftarrow_{\$} \mathsf{Perm}(n) : \forall (x, y) \in \tau, P(P(x) \oplus x) = y\right]. \tag{1.13}$$

Soit

$$r = |Y| \tag{1.14}$$

le nombre de réponses distinctes de l'oracle apparaissant dans la transcription et

$$s = |\{(x, y) \in \tau : \forall (x', y') \in \tau \setminus \{(x, y)\}, y' \neq y\}| \tag{1.15}$$

le nombre de requêtes qui ne sont pas en collision.

Comme nous l'avons observé, pour minorer la probabilité d'obtenir τ dans le monde réel, nous devons minorer le nombre de permutations P telles que

$$\forall (x, y) \in \tau,\ P(P(x) \oplus x) = y. \tag{1.16}$$

Le comptage du nombre de solutions est rendu plus difficile que dans la section précédente à cause du fait que ces équations de sont pas « indépendantes ». Par exemple, s'il existe deux requêtes (x, y) et (x', y') dans τ tels que $P(x) \oplus x = x'$, alors on doit nécessairement avoir $P(x') = y$. De même, si $P(x) = y'$, alors on doit avoir $P(x') \oplus x' = x$. Nous pourrions nous contenter de compter les permutations P telles que, pour toute requête $(x, y) \in \tau$, $P(x) \oplus x \notin X \cup Y$, ce qui éviterait toute collision interne. Malheureusement, une telle simplification ne peut mener qu'à une borne de sécurité au niveau de la borne des anniversaires. Ainsi, pour dépasser cette limite, nous aurons besoin d'être plus précis. Comme nous allons le voir, il sera suffisant de considérer les permutations P telles que $P(x) \oplus x = x'$ pour exactement t paires de requêtes $((x, y), (x', y'))$ qui ne sont pas en collision, lorsque t varie dans un intervalle suffisamment étendu. Il faut cependant être attentif, lors du choix de ces t paires de requêtes, à ne pas créer de contrainte incompatible avec une autre requête présente dans la transcription. Dans ce but, nous introduisons la définition suivante.

Définition 4. *Une combinaison de t paires deux à deux distinctes de requêtes qui ne sont pas en collision*

$$\Sigma = \{((x_1, y_1), (x'_1, y'_1)), \ldots, ((x_t, y_t), (x'_t, y'_t))\}$$

est dite bonne *si les conditions suivantes sont vérifiées :*

(a) pour tout $i \in \{1, \ldots, t\}$, $y_i \oplus x'_i \notin X$;

(b) pour tout $i \in \{1, \ldots, t\}$, $x_i \oplus x'_i \notin Y$;

(c) les valeurs $y_i \oplus x'_i$, $i \in \{1, \ldots, t\}$, sont deux à deux distinctes ;

(d) les valeurs $x_i \oplus x'_i$, $i \in \{1, \ldots, t\}$, sont deux à deux distinctes.

Notons à présent

$$M = \frac{q}{2^{n/3}}. \tag{1.17}$$

Alors on a le lemme suivant, qui montre que le nombre de bonnes combinaisons Σ est proche de $(s)_{2t}/t!$, le nombre total de combinaisons de t paires de requêtes qui ne sont pas en collision. La preuve de ce lemme est donnée en annexe B.3.

Lemme 10. *Fixons un entier naturel t tel que $0 \leq t \leq M$. Alors le nombre $N(t)$ de bonnes combinaisons Σ de t paires de requêtes qui ne sont pas en collision vérifie*

$$N(t) \geq \frac{(s)_{2t}}{t!} \left(1 - \frac{8\sqrt{3nq}}{2^{n/3}} - \frac{12q}{2^{2n/3}}\right).$$

(On rappelle que s correspond au nombre de requêtes qui ne sont pas en collision dans τ.)

À présent, fixons un entier naturel t tel que $0 \leq t \leq M$ ainsi qu'une bonne combinaison de t paires de requêtes qui ne sont pas en collision, notée

$$\Sigma = \{((x_1, y_1), (x'_1, y'_1)), \ldots, ((x_t, y_t), (x'_t, y'_t))\}.$$

Nous allons minorer le nombre de permutations P qui vérifient (1.16) et telles que, pour tout $i \in \{1, \ldots, t\}$, $P(x_i) \oplus x_i = x'_i$. Notons qu'une telle permutation vérifie (1.16) pour les $2t$ requêtes apparaissant dans Σ si et seulement si

$$\forall i \in \{1, \ldots, t\}, \begin{cases} P(x_i) = x_i \oplus x'_i \\ P(x'_i) = y_i \\ P(y_i \oplus x'_i) = y'_i. \end{cases} \tag{1.18}$$

Cet ensemble de $3t$ équation est « vérifiable », dans le sens où toutes les entrées, respectivement toutes les sorties, sont deux à deux distinctes d'après les conditions (a) et (c), respectivement (b) et (d), qui caractérisent une bonne combinaison Σ (mais aussi d'après le fait que les valeurs x_i sont deux à deux distinctes par supposition, et le fait que les y_i sont deux à deux distincts dans le cas de requêtes qui ne sont pas en collision).

Dans la suite de cette section, notons

$$X' = X \cup \{y_i \oplus x'_i : i \in \{1, \ldots, t\}\}$$
$$Y' = Y \cup \{x_i \oplus x'_i : i \in \{1, \ldots, t\}\}.$$

Remarquons également que $|X'| = q + t$ et $|Y'| = r + t$.

Il reste à considérer les $q - 2t$ requêtes $(u, v) \in \tau$ qui n'apparaissent pas dans Σ. Soient $q' = q - 2t$ le nombre de ces requêtes, $r' = r - 2t$ le nombre de réponses distinctes de l'oracle qui apparaissent dans ces requêtes, et $s' = s - 2t$ le nombre de ces requêtes qui ne sont pas en collision. Nous allons regrouper ces requêtes restantes de sorte que toutes les requêtes partageant la même réponse soient consécutives, et les notons de la façon suivante

$$\tau' = \Big((u_{1,1}, v_1), \ldots, (u_{1,q_1}, v_1), \\ \ldots, \\ (u_{r',1}, v_{r'}), \ldots, (u_{r',q_{r'}}, v_{r'})\Big),$$

où $v_1, \ldots, v_{r'}$ sont deux à deux distincts et $\sum_{i=1}^{r'} q_i = q'$. Afin de faciliter les calculs, nous supposerons également que nous avons ordonné les requêtes de sorte que les requêtes qui ne sont pas en collision apparaissent en premier, c'est-à-dire que $q_i = 1$ pour $i \in \{1, \ldots, s'\}$ et $q_i > 1$ pour $i \in \{s' + 1, \ldots, r'\}$. Notons que, puisque τ est une bonne transcription, on a

$$\sum_{i=s'+1}^{r'} q_i \leq \sqrt{q} \tag{1.19}$$

car sinon la condition (i) définissant une mauvaise transcription serait vérifiée.

Notre but, à présent, est de minorer le nombre de permutations P qui, en plus de vérifier les équations (1.18), vérifient également

$$\forall (u, v) \in \tau',\ P(P(u) \oplus u) = v. \tag{1.20}$$

Dans ce but, comme dans la section précédente, nous allons considérer toutes les valeurs « intermédiaires » possibles $z_i = P^{-1}(v_i)$. Formellement, nous allons avoir besoin de la définition ci-dessous.

Définition 5. *Un uplet de r' valeurs $\mathbf{z} = (z_1, \ldots, z_{r'})$ est dit* bon *si tous les z_i sont deux à deux distincts et à l'extérieur de X', et toutes les valeurs $z_i \oplus u_{i,j}$ pour $i \in \{1, \ldots, r'\}$ et $j \in \{1, \ldots, q_i\}$ sont deux à deux distinctes et en dehors de Y'.*

Remarquons que, pour tout bon uplet $\mathbf{z} = (z_1, \ldots, z'_r)$, le système d'équations

$$\begin{cases} \forall i \in \{1, \ldots, r'\},\ \forall j \in \{1, \ldots, q_i\},\ P(u_{i,j}) = z_i \oplus u_{i,j} \\ \forall i \in \{1, \ldots, r'\},\ P(z_i) = v_i \end{cases} \tag{1.21}$$

est « vérifiable » et « compatible » avec les équations (1.18) dans le sens où toutes les entrées, respectivement les sorties qui apparaissent dans les équations (1.18) et (1.21) sont deux à deux distinctes par définition d'une bonne combinaison Σ et d'un bon uplet $\mathbf{z}$. De plus, une permutation P qui vérifie les équations (1.21) est telle que $P(P(u) \oplus u) = v$ pour tout $(u, v) \in \tau'$.

Lemme 11. *Fixons t et Σ comme ci-dessus. Alors le nombre $N'(t)$ de bons uplets $\mathbf{z}$ vérifie*

$$N'(t) \geq \prod_{i=0}^{r'-1} \left(2^n - q - t - i - q_{i+1} \left(r + t + \sum_{j=1}^{i} q_j \right) \right). \tag{1.22}$$

La preuve de ce lemme peut être trouvée dans l'annexe B.4.

Nous pouvons à présent réunir tous ces éléments pour effectuer le comptage final. Pour tout entier naturel t tel que $0 \leq t \leq M$, et chaque choix possible de bonne combinaison Σ de t paires de requêtes qui ne sont pas en collision, et chaque choix possible de bon uplet $\mathbf{z}$, la probabilité qu'une permutation P choisie uniformément aléatoirement vérifie les équations (1.18) et (1.21) (ce qui entraîne qu'elle vérifie le système d'équations (1.16)) vaut exactement

$$\frac{1}{(2^n)_{q+r-t}}.$$

En effet, il y a exactement $3t$ équations dans le système (1.18) et exactement

$$q' + r' = (q - 2t) + (r - 2t) = q + r - 4t$$

équations dans le système (1.21), ainsi $q + r - t$ équations au total doivent être vérifiées, et celles-ci sont « compatibles » par définition d'une bonne combinaison Σ et d'un bon uplet $\mathbf{z}$. En sommant sur les choix possibles pour t, Σ, and $\mathbf{z}$, on obtient que la probabilité d'obtenir la transcription τ dans le monde réel vérifie

$$\Pr\left[T_{\text{re}} = \tau\right] \geq \sum_{0 \leq t \leq M} \frac{N(t)N'(t)}{(2^n)_{q+r-t}}.$$

Puisque la probabilité d'obtenir τ dans le monde idéal vaut simplement $1/(2^n)^q$, le rapport entre ces deux probabilités vérifie

$$\rho \stackrel{\text{def}}{=} \frac{\Pr\left[T_{\text{re}} = \tau\right]}{\Pr\left[T_{\text{id}} = \tau\right]} \geq \sum_{0 \leq t \leq M} \frac{(2^n)^q N(t) N'(t)}{(2^n)_{q+r-t}}. \tag{1.23}$$

En réunissant les lemmes 10 et 11, puis en utilisant plusieurs astuces techniques, il est possible de prouver le lemme suivant.

Lemme 12. *Supposons* $36 \leq q \leq 2^n/6$. *Alors, pour toute bonne transcription* τ, *on a*

$$\rho = \frac{\Pr\left[T_{\text{re}} = \tau\right]}{\Pr\left[T_{\text{id}} = \tau\right]} \geq 1 - \frac{8\sqrt{3nq}}{2^{n/3}} - \frac{50q}{2^{2n/3}}.$$

La preuve de ce lemme est assez calculatoire et est donc reportée à l'annexe B.5.

1.7.4 Conclusion

Nous pouvons à présent achever la preuve du théorème 3. Pour ce faire, il suffit de combiner les lemmes 9 et 12 avec le lemme 1. On obtient ainsi, comme annoncé, le fait que l'avantage du distingueur $\mathcal{D}$ est inférieur à

$$\frac{52q}{2^{2n/3}} + \frac{8\sqrt{3nq}}{2^{n/3}}.$$

Chapitre 2

Le schéma d'Even-Mansour paramétrable

2.1 Le schéma d'Even-Mansour

De nombreux algorithmes de chiffrement par blocs font partie de la famille des algorithmes de chiffrement à clé alternée : on peut notamment citer AES [DR02], PRESENT [BKL+07] ou encore LED [GPPR11]. Les algorithmes de cette famille sont construits selon une structure commune très simple : étant donné r permutations publiques $P_1, \ldots, P_r$ de l'ensemble des messages de n bits, on calcule le chiffré y d'un message clair x de n bits de la façon suivante

$$y = k_r \oplus P_r(k_{r-1} \oplus P_{r-1}(\cdots P_2(k_1 \oplus P_1(k_0 \oplus x)) \cdots)), \tag{2.1}$$

où $(k_0, \ldots, k_r)$ est une suite de sous-clés de n bits, généralement dérivées d'une clé maître. Il est alors naturel de se demander si cette structure simple est saine, c'est-à-dire s'il existe des attaques contre ce type de schéma qui fonctionnent quel que soit le choix de permutations $P_1, .., P_r$. Afin d'étudier la résistance des algorithmes de chiffrement à clé alternée à ce type d'attaques, appelées attaques génériques, les cryptographes ont construit une abstraction de ce type de schéma en modélisant les permutations $P_1, \ldots, P_r$ par des oracles de permutation aléatoires et indépendants et en voyant les sous-clés $k_0, \ldots, k_r$ comme $r+1$ clés de n bits choisies indépendamment avec une distribution uniforme. En plus de pouvoir effectuer des requêtes de chiffrement ou de déchiffrement, l'adversaire a le pouvoir d'effectuer des requêtes en boîte noire aux permutations internes (les requêtes peuvent être dans le sens direct ou dans le sens inverse). Il s'agit d'un modèle très fort qui permet de donner des preuves d'indistinguabilité face à des adversaires ayant une puissance de calcul illimitée, qui indiquent que, pour être efficace, une attaque contre ce type d'algorithmes doit exploiter les caractéristiques des permutations internes.

Historiquement, le premier résultat dans ce modèle a été obtenu pour un tour ($r = 1$) par Even et Mansour [EM97] et c'est pour cette raison que, par la suite, cette abstraction des algorithmes de chiffrement à clé alternée a été nommée schéma d'Even-Mansour. Dans cet article initial, les auteurs ont démontré que l'algorithme de chiffrement associant un clair x à son chiffré $k_1 \oplus P_1(k_0 \oplus x)$, où k_0 et k_1 sont

deux clés de n bits choisies indépendamment avec une distribution uniforme et P_1 est une permutation publique uniformément aléatoire, est sûre tant que le nombre de requêtes autorisées à l'adversaire est petit devant $2^{n/2}$. Bogdanov *et al.* [BKL+12] ont ensuite démontré que, lorsque le nombre r de tours est supérieur à 2, la sécurité est garantie tant que le nombre de requêtes de l'adversaire est petit devant $2^{2n/3}$. Ils ont également émis la conjecture que, dans le cas général, la sécurité devrait être garantie pour un nombre quelconque de tours r tant que le nombre de requêtes est inférieur à $2^{\frac{rn}{r+1}}$, tout en exhibant un distingueur correspondant à cette borne. Steinberger [Ste12] a ensuite prouvé que cette conjecture était vraie pour $r = 3$. En 2012, Lampe *et al.* [LPS12] ont démontré, à l'aide de la technique du couplage, que la sécurité du schéma d'Even-Mansour à r tours, pour tout entier r pair, est garantie tant que le nombre de requêtes est inférieur à $2^{\frac{rn}{r+2}}$, ce qui constitue la première borne de sécurité qui s'améliore en fonction du nombre de tours en tendant asymptotiquement vers la sécurité optimale. Moins de deux ans plus tard, Chen et Steinberger [CS14] ont finalement prouvé à l'aide de la technique des coefficients H que la conjecture de Bogdanov *et al.* était vraie dans le cas général.

Récemment, une nouvelle voie de recherche s'est ouverte, qui propose de généraliser le schéma d'Even-Mansour. Des versions minimalisées de ce schéma ont notamment été étudiées par Chen *et al.* [CLL+14] qui ont prouvé que, à un facteur logarithmique près, le schéma d'Even-Mansour à 2 tours reste sûr lorsque le nombre de requêtes de l'adversaire est inférieur à $2^{2n/3}$, même lorsque les permutations P_1 et P_2 sont identiques et que les sous-clés k_0, k_1, k_2 sont dérivées d'une même clé de n bits k, pour une certaine famille de fonctions linéaires de dérivation de clés. Ce résultat présente l'avantage de modéliser de manière plus réaliste les algorithmes de chiffrement à clés alternées, notamment en réduisant le nombre de bits de clé et en ne modélisant pas les permutations de tour comme deux permutations indépendantes. L'extension de ce résultat à un nombre plus élevé de tours reste un problème ouvert très difficile. Les résultats de ce manuscrit se concentrent sur un autre type de généralisation, celle du schéma d'Even-Mansour paramétrable. L'objectif visé dans ces travaux est de donner une structure résistant aux attaques génériques à partir de laquelle il est possible de construire des algorithmes de chiffrement paramétrables, qui étendent la notion d'algorithme de chiffrement par blocs.

2.2 Les algorithmes de chiffrement par blocs paramétrables

Dans le chapitre précédent, les algorithmes de chiffrement par blocs ayant $\mathcal{K}$ pour ensemble de clés et $\mathcal{M}$ pour ensemble de messages ont été définis comme une famille de permutations de $\mathcal{M}$ paramétrée par la clé $\mathbf{k} \in \mathcal{K}$. Un algorithme de chiffrement par bloc *paramétrable* admet un paramètre additionnel $\mathbf{t} \in \mathcal{T}$, éventuellement public, appelé un *tweak* en anglais. Son rôle est d'apporter une variabilité au chiffrement qui soit directement intégrée à l'algorithme, à la manière d'un vecteur d'initialisation ou d'un nonce pour un mode d'opération par exemple. Certains algorithmes de chiffrement intègrent nativement cette fonctionnalité comme le Hasty Pudding

Cipher [Sch98], Mercy [Cro00], ou Threefish (l'algorithme de chiffrement par blocs utilisé dans la famille de fonctions de hachage Skein [FLS+10]). La syntaxe et les notions de sécurité des algorithmes de chiffrement par blocs paramétrables ont été formellement définies dans un article initial de Liskov, Rivest et Wagner [LRW02]. Il est notamment attendu d'un tel algorithme qu'il soit indistinguable d'une permutation paramétrable (c'est-à-dire une famille de permutations aléatoires et paramétrées par l'ensemble des *tweaks*). Depuis lors, cette primitive a trouvé plusieurs applications comme les modes de chiffrement (paramétrables) préservant la longueur [HR03, HR04], le chiffrement en ligne [RZ11, ABL+13], le chiffrement authentifié [LRW02, RBB03, Rog04] ainsi que le chiffrement de disque dur.

2.2.1 Les constructions génériques

Nous allons présenter différentes constructions génériques d'un algorithme de chiffrement paramétrable à partir d'un algorithme de chiffrement par bloc standard. Soient E un algorithme de chiffrement par blocs arbitraire, dont la sécurité est établie, d'ensemble de clés et de messages $\{0,1\}^n$ et $\mathcal{H}$ une famille de fonctions ε-AXU d'un ensemble $\mathcal{T}$ vers $\{0,1\}^n$, c'est-à-dire que $\mathcal{H}$ vérifie la condition suivante :

$$\forall t \in \mathcal{T},\, \forall t' \in \mathcal{T} \setminus \{t\},\, \forall y \in \{0,1\}^n,\, \Pr\left[h \leftarrow_{\$} \mathcal{H} \,:\, h(t) \oplus h(t') = y\right] \leq \varepsilon.$$

Dans leur article fondateur, Liskov *et al.* [LRW02] ont proposé deux constructions génériques différentes et obtenu une preuve de sécurité jusqu'à la borne des anniversaires pour chaque construction, i.e., quand l'adversaire a droit à au plus $2^{n/2}$ requêtes aux oracles de chiffrement ou de déchiffrement. Ces deux constructions sont les suivantes :

$$\begin{aligned}\mathrm{LRW}^1_k(t,m) =& E_k(t \oplus E_k(m)),\\ \mathrm{LRW}^2_{h,k}(t,m) =& h(t) \oplus E_k(m \oplus h(t)),\end{aligned}$$

pour tout $t \in \mathcal{T}$, $k \in \{0,1\}^n$, $m \in \{0,1\}^n$ et tout $h \in \mathcal{H}$. Notons que, pour LRW^1, on a nécessairement $\mathcal{T} = \{0,1\}^n$. La simplicité de ces constructions les rend très utiles en pratique. Toutefois, leur sécurité est limitée à la borne des anniversaires, ce qui peut s'avérer trop faible dans certaines applications.

Il existe de nombreuses autres stratégies de conception « en boîte noire ». On peut notamment citer des constructions comme XEX [Rog04] et ses variantes [Min06, CS06] qui sont liées à la première proposition de Liskov *et al.*, et souffrent de la même limitation de sécurité à la borne des anniversaires. Plus récemment, un certain nombre de constructions bénéficiant d'une sécurité au-delà de la borne des anniversaires ont été publiées. Par exemple, on peut remarquer que la construction LRW^2 peut être itérée avec des clés indépendantes, ce qui permet d'en augmenter le niveau de sécurité au-delà de la borne des anniversaires et en le faisant tendre asymptotiquement vers la sécurité optimale [LST12, LS14, Pro14]. Cependant, à mesure que la cascade augmente en longueur, la quantité de calculs nécessaire à l'évaluation du schéma augmente également, ainsi que la taille de la clé, ce qui rend cette construction peu pratique. Il existe également d'autres constructions qui n'ont pas besoin d'être

itérées, mais offrent tout de même des garanties de sécurité au-delà de la borne des anniversaires, comme par exemple la construction de Minematsu [Min09], notée Min, ou celles Mennink [Men15], notées $\tilde{F}[1]$ et $\tilde{F}[2]$:

$$\begin{aligned}
\mathrm{Min}_k(t,m) =& E(E(k, \underbrace{t}_{\theta \text{ bits}} \,||\, \underbrace{0\ldots0}_{n-\theta \text{ bits}}), m) && \text{pour tout } (t,k,m) \in \{0,1\}^\theta \times (\{0,1\}^n)^2 \\
\tilde{F}[1]_k(t,m) =& E_{k\oplus t}(m \oplus k \otimes t) \oplus k \otimes t && \text{pour tout } (k,t,m) \in (\{0,1\}^n)^3 \\
\tilde{F}[2]_k(t,m) =& E_{k\oplus t}(m \oplus E_k(t)) \oplus E_k(t) && \text{pour tout } (k,t,m) \in (\{0,1\}^n)^3
\end{aligned}$$

où $\otimes$ désigne le produit sur le corps fini à 2^n éléments, pour un polynôme irréductible arbitraire fixé. Une condition nécessaire pour la sécurité de la construction de Minematsu est que le nombre de requêtes de l'adversaire soit petit devant $2^{n-\theta}$: ainsi, la construction est d'autant plus sûre que l'espace de *tweaks* est restreint. Les constructions de Mennink sont sûres tant que le nombre de requêtes de l'adversaire est petit devant $2^{2n/3}$, respectivement 2^n. Toutes ces constructions ont une preuve de sécurité dans le modèle standard (c'est-à-dire qu'elles supposent uniquement que l'algorithme de chiffrement par blocs sous-jacent fournit une permutation pseudo-aléatoire lorsque la clé est choisie uniformément aléatoirement), à l'exception des constructions de Mennink qui sont analysées dans le modèle de l'algorithme de chiffrement idéal (c'est-à-dire que l'on modélise E comme un algorithme de chiffrement par blocs tiré uniformément aléatoirement auquel l'adversaire a accès en boîte noire). Il est facile de voir que, puisque l'algorithme de chiffrement E est utilisé avec la clé $k \oplus t$, le simple fait que E_k pour une clé aléatoire k soit indistinguable d'une permutation aléatoire n'est pas suffisant ; ainsi E doit également résister à une certaine classe d'attaques à clés reliées [BK03], pour laquelle l'attaquant est capable de XORer des constantes de son choix à la clé utilisée par l'algorithme de chiffrement avec lequel il interagit. Malheureusement, aucune de ces constructions génériques ayant une sécurité au-delà de la borne des anniversaires ne peut vraiment être considérée comme applicable en pratique (bien que certaines d'entre elles puissent s'en approcher [Men15]). Ainsi, il est naturel d'étudier comment construire un algorithme de chiffrement par blocs à partir d'une primitive de plus bas niveau qu'un algorithme de chiffrement par blocs conventionnel, par exemple une permutation ou une fonction pseudo-aléatoire.

2.2.2 Construction d'algorithmes de chiffrement par blocs nativement paramétrables

Les premières structures permettant la construction d'algorithmes de chiffrement nativement paramétrables ont été étudiées par Goldenberg *et al.* [GHL+07] qui ont cherché à inclure un *tweak* dans un réseau de Feistel. Ce travail a été étendu aux réseaux de Feistel généralisés par Mitsuda et Iwata [MI08]. Nous avons mené une étude similaire pour la deuxième grande classe d'algorithmes de chiffrement par blocs, les algorithmes de chiffrement à clé alternée, modélisés comme nous l'avons vu par la construction d'Even-Mansour [CS15b, CLS15, CS15a] et c'est à ces travaux qu'est consacré le reste de ce manuscrit.

Afin d'incorporer un *tweak* $\mathbf{t}$ aux schémas d'Even-Mansour, il est tentant de généraliser (2.1) en remplaçant les clés de tour k_i par une fonction $f_i(\mathbf{k}, \mathbf{t})$ de la clé maître $\mathbf{k}$ *et* du *tweak* $\mathbf{t}$ (voir figure 2.1). Nous appellerons une telle construction le schéma d'Even-Mansour paramétrable (TEM). Nous attirons l'attention sur le fait que cette dénomination (en anglais *Tweakable Even-Mansour*) a été précédemment utilisée par les concepteurs de Minalpher [STA+14], un candidat à la compétition CAESAR, pour désigner une variante basée sur des permutations de la construction XEX de Rogaway [Rog04], c'est-à-dire un schéma d'Even-Mansour à un tour dans lequel les fonctions de dérivation de clé f_0 et f_1 appliquées à $(\mathbf{k}, \mathbf{t})$ dépendent de la permutation P_1. De telles constructions ont également été étudiées par Mennink [Men16] et Granger *et al* [GJMN16]. Dans ce manuscrit, nous considérons des constructions pour lesquelles les permutations internes sont indépendantes des fonctions de dérivation de clés. Signalons que le cas particulier de la construction à un tour dont nous allons prouver la sécurité dans la section 3.2 a déjà été étudié par Kurosawa [Kur10], ce dont nous n'avions pas connaissance lors de la rédaction de notre article publié à CRYPTO 2015.

L'idée de faire dépendre les clés de tour du *tweak* a également été introduite par Jean *et al.* [JNP14] dans leur construction TWEAKEY. Ils y présentent des critères concrets de conception d'algorithmes de chiffrement par blocs paramétrables sur le modèle des algorithmes de chiffrement à clé alternée. Leur approche est toutefois différente de la nôtre puisque leurs arguments sont fondés sur l'étude d'attaques tandis que nous nous concentrons sur des preuves mathématiques de sécurité.

2.3 Notations et transcriptions

2.3.1 Algorithmes de chiffrement par blocs paramétrables

Un *algorithme de chiffrement par blocs paramétrable* ayant pour espace de clés $\mathcal{K}$, espace de *tweaks* $\mathcal{T}$, et espace de message $\mathcal{M}$ est une fonction $\widetilde{E} : \mathcal{K} \times \mathcal{T} \times \mathcal{M} \to \mathcal{M}$ telle que, pour toute clé $k \in \mathcal{K}$ et tout *tweak* $t \in \mathcal{T}$, $x \mapsto \widetilde{E}(k, t, x)$ est une permutation inversible de $\mathcal{M}$. Notons $\mathsf{TBC}(\mathcal{K}, \mathcal{T}, n)$ l'ensemble des algorithmes de chiffrement par blocs paramétrables avec espace de clé $\mathcal{K}$, espace de *tweak* $\mathcal{T}$, et espace de message $\{0,1\}^n$. Une *permutation paramétrable* dont l'espace de *tweaks* est $\mathcal{T}$ et l'espace de messages est $\mathcal{M}$ est une fonction $\widetilde{P} : \mathcal{T} \times \mathcal{M} \to \mathcal{M}$ telle que, pour tout *tweak* $t \in \mathcal{T}$, $x \mapsto \widetilde{P}(t, x)$ est une permutation de $\mathcal{M}$. On notera $\mathsf{TP}(\mathcal{T}, n)$ l'ensemble de toutes les permutations paramétrables ayant $\mathcal{T}$ pour espace de *tweaks* et $\{0,1\}^n$ pour espace de messages.

2.3.2 Schéma d'Even-Mansour paramétrable

Fixons des entiers naturels $n, r \geq 1$. Soient $\mathcal{K}$ et $\mathcal{T}$ deux ensembles, et soit $\mathbf{f} = (f_0, \ldots, f_r)$ un $(r+1)$-uplet de fonctions de $\mathcal{K} \times \mathcal{T}$ vers $\{0,1\}^n$. Le schéma d'Even-Mansour paramétrable à r tours $\mathsf{TEM}[n, r, \mathbf{f}]$ décrit, à partir d'un r-uplet $\mathbf{P} = (P_1, \ldots, P_r)$ de permutations de $\{0,1\}^n$, un algorithme de chiffrement par blocs paramétrable dont l'ensemble des clés est $\mathcal{K}$, celui des *tweaks* $\mathcal{T}$, et celui des messages

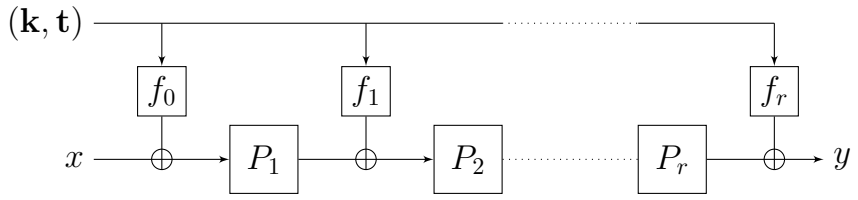

FIGURE 2.1 – Le schéma d'Even-Mansour à r tours utilisant sur un r-uplet de permutations publiques $(P_1, \ldots, P_r)$.

$\{0,1\}^n$, simplement noté $\mathsf{TEM}^{\mathbf{P}}$ dans la suite (les paramètres $[n, r, \mathbf{f}]$ seront toujours clairs en fonction du contexte) qui, à partir d'une clé $\mathbf{k} \in \mathcal{K}$, d'un *tweak* $\mathbf{t} \in \mathcal{T}$, envoie un texte clair $x \in \{0,1\}^n$ sur un chiffré défini par (voir la figure 2.1) :

$$\mathsf{TEM}^{\mathbf{P}}(\mathbf{k}, \mathbf{t}, x) = f_r(\mathbf{k}, \mathbf{t}) \oplus P_r(f_{r-1}(\mathbf{k}, \mathbf{t}) \oplus P_{r-1}(\cdots P_1(f_0(\mathbf{k}, \mathbf{t}) \oplus x) \cdots)).$$

Nous noterons $\mathsf{TEM}^{\mathbf{P}}_{\mathbf{k}}$ la fonction prenant en entrée $(\mathbf{t}, x) \in \mathcal{T} \times \{0,1\}^n$ et renvoyant $\mathsf{TEM}^{\mathbf{P}}(\mathbf{k}, \mathbf{t}, x)$.

Nous nous focaliserons la plupart du temps sur le cas où $\mathcal{K} = (\{0,1\}^n)^a$ et $\mathcal{T} = (\{0,1\}^n)^b$ pour des entiers naturels $a, b \geq 1$. Dans ce contexte, nous noterons $\mathbf{k} = (k_0, \ldots, k_{a-1})$ et $\mathbf{t} = (t_0, \ldots, t_{b-1})$, les k_i et les t_j étant des chaînes de caractères de n bits, ou simplement $\mathbf{k} = k$, respectivement $\mathbf{t} = t$ quand $a = 1$, respectivement $b = 1$. Quand les f_i sont linéaires sur $(\{0,1\}^n)^{a+b}$, nous dirons que la construction a un *mixage linéaire du tweak et de la clé*.

2.3.3 Description des transcriptions

Soient q_c et q_p deux entiers naturels et soit r un entier naturel non-nul. On rappelle que le schéma d'Even-Mansour paramétrable est étudié dans le modèle de la permutation aléatoire : les permutations de tour $P_1, \ldots, P_r$ sont modélisées par des permutations uniformément aléatoires publiques. Dans ce contexte, un distingueur $\mathcal{D}$ est, comme précédemment, un algorithme déterministe dont la puissance de calcul n'est pas bornée et qui interagit avec un uplet de $r+1$ oracles noté $(\widetilde{P}_0, P_1, \ldots, P_r)$. Dans le monde réel, l'oracle de construction $\widetilde{P}_0$ est $\mathsf{TEM}^{\mathbf{P}}_{\mathbf{k}}$ où $\mathbf{P} = (P_1, \ldots, P_r)$ et $\mathbf{k}$ est aléatoire, alors que dans le monde idéal il s'agit d'une permutation paramétrable uniformément aléatoire indépendante de $(P_1, \ldots, P_r)$. À partir de l'interaction de $\mathcal{D}$ avec ces oracles, on définit la *transcription des requêtes* $(\mathcal{Q}_C, \mathcal{Q}_{P_1}, \ldots, \mathcal{Q}_{P_r})$ de l'attaque comme suit. La liste $\mathcal{Q}_C$ enregistre les requêtes à l'oracle de construction : si $\mathcal{D}$ effectue une requête directe $(\mathbf{t}, x)$ à l'oracle de construction $\widetilde{P}_0$ et reçoit la réponse y, ou une requête inverse $(\mathbf{t}, y)$ et reçoit la réponse x, alors le triplet $(\mathbf{t}, x, y) \in \mathcal{T} \times \{0,1\}^n \times \{0,1\}^n$ est ajouté à $\mathcal{Q}_C$. De même, pour $1 \leq i \leq r$, $\mathcal{Q}_{P_i}$ contient toutes les paires $(u, v) \in \{0,1\}^n \times \{0,1\}^n$ telles que $\mathcal{D}$ a effectué soit une requête directe u à la permutation P_i et reçu pour réponse v, soit une requête inverse v et reçu pour réponse u. On remarquera que les requêtes sont enregistrées de façon non-ordonnée et en oubliant la direction de la requête, mais grâce à l'hypothèse que le distingueur est déterministe, l'interaction de $\mathcal{D}$ avec ses oracles peut être totalement

reconstruite à partir de la transcription des requêtes (voir par exemple [CS14] pour plus de détails). Notons également que, puisqu'on a supposé que $\mathcal{D}$ n'effectue jamais de requêtes inutiles, chaque requête à l'oracle de construction résulte en un nouveau triplet, distinct des précédents, dans $\mathcal{Q}_C$, et chaque requête à P_i résulte en une paire différente dans $\mathcal{Q}_{P_i}$. De plus, puisque nous avons supposé que le distingueur effectue toujours le nombre maximal de requêtes autorisées à chaque oracle, on a $|\mathcal{Q}_C| = q_c$ et $|\mathcal{Q}_{P_i}| = q_p$ pour $1 \leq i \leq r$. Dans la suite de ce manuscrit, on notera m le nombre de *tweaks* distincts apparaissant dans $\mathcal{Q}_C$, et q_i le nombre de requêtes pour le $i$ème *tweak*, $1 \leq i \leq m$, en ordonnant les *tweaks* de façon arbitraire. Remarquons que l'on a toujours $\sum_{i=1}^{m} q_i = q_c$, bien que m puisse dépendre des réponses reçues des oracles.

Une transcription des requêtes est dite *atteignable* (par rapport à un distingueur fixé $\mathcal{D}$) s'il existe des oracles $(\widetilde{P}_0, \mathbf{P})$ tels que l'interaction de $\mathcal{D}$ avec $(\widetilde{P}_0, \mathbf{P})$ résulte en cette transcription (en d'autres termes, la probabilité d'obtenir cette transcription dans le monde idéal est non-nulle). De plus, dans le but d'avoir une définition simple des mauvaises transcriptions, la clé utilisée $\mathbf{k}$ est révélée à l'adversaire à la fin de l'expérience si nous sommes dans le monde réel, alors que, dans le monde idéal, une clé factice $\mathbf{k} \leftarrow_{\$} \mathcal{K}$ est simplement tirée uniformément aléatoirement indépendamment des réponses de l'oracle $\widetilde{P}_0$ (il n'y a évidemment aucune perte de généralité puisque cette information ne peut qu'aider le distingueur et augmenter son avantage). Pour conclure, une transcription τ est un uplet $\tau = (\mathcal{Q}_C, \mathcal{Q}_{P_1}, \ldots, \mathcal{Q}_{P_r}, \mathbf{k})$, et une transcription est dite atteignable si la transcription des requêtes correspondante $(\mathcal{Q}_C, \mathcal{Q}_{P_1}, \ldots, \mathcal{Q}_{P_r})$ est atteignable. On note Θ l'ensemble des transcriptions atteignables. Par la suite, on notera T_{re}, respectivement T_{id}, la distribution de probabilité de la transcription τ induite par le monde réel, respectivement le monde idéal (notons que ces deux distributions de probabilité dépendent du distingueur). Par extension, on utilise la même notation pour noter une variable aléatoire distribuée selon chaque distribution.

2.3.4 Quelques observations utiles

Nous achevons ce chapitre avec quelques observations préliminaires. Tout d'abord, introduisons quelques notations supplémentaires. Étant donné une transcription des requêtes à un oracle de permutation $\mathcal{Q}$ et une permutation P, on dit que P *prolonge* $\mathcal{Q}$, évènement noté $P \vdash \mathcal{Q}$, si $P(u) = v$ pour tout $(u, v) \in \mathcal{Q}$. Par extension, étant donné un uplet de transcriptions de requêtes à des oracles de permutation $\mathcal{Q}_{\mathbf{P}} = (\mathcal{Q}_{P_1}, \ldots, \mathcal{Q}_{P_r})$ et un uplet de permutations $\mathbf{P} = (P_1, \ldots, P_r)$, on dit que $\mathbf{P}$ prolonge $\mathcal{Q}_{\mathbf{P}}$, évènement noté $\mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}}$, si $P_i \vdash \mathcal{Q}_{P_i}$ pour chaque $i = 1, \ldots, r$. Notons que, pour une telle transcription de taille q_p, on a

$$\Pr[P \leftarrow_{\$} \mathsf{P}(n) : P \vdash \mathcal{Q}] = \frac{1}{(N)_{q_p}}. \tag{2.2}$$

De même, étant donné la transcription de l'interaction avec un oracle de construction $\widetilde{\mathcal{Q}}$ et une permutation paramétrable $\widetilde{P}$, on dit que $\widetilde{P}$ *prolonge* $\widetilde{\mathcal{Q}}$, évènement noté $\widetilde{P} \vdash \widetilde{\mathcal{Q}}$, si $\widetilde{P}(t, x) = y$ pour tout $(t, x, y) \in \widetilde{\mathcal{Q}}$. Pour une telle transcription $\widetilde{\mathcal{Q}}$ dans laquelle apparaissent m *tweaks* distincts et q_i requêtes correspondant au i-ème *tweak*,

on a

$$\Pr[\widetilde{P} \leftarrow_{\$} \mathsf{TP}(\mathcal{T}, n) : \widetilde{P} \vdash \widetilde{\mathcal{Q}}] = \prod_{i=1}^{m} \frac{1}{(N)_{q_i}}. \tag{2.3}$$

Il est clair que l'interaction d'un distingueur $\mathcal{D}$ avec les oracles $(\widetilde{P}_0, P_1, \ldots, P_r)$ aboutit à une transcription atteignable des requêtes $(\mathcal{Q}_C, \mathcal{Q}_{\mathbf{P}})$ où $\mathcal{Q}_{\mathbf{P}} = (\mathcal{Q}_{P_1}, \ldots, \mathcal{Q}_{P_r})$ *si et seulement si* $\widetilde{P}_0 \vdash \mathcal{Q}_C$ et $P_i \vdash \mathcal{Q}_{P_i}$ pour $1 \leq i \leq r$. Dans le monde idéal, la clé $\mathbf{k}$, les permutations $P_1, \ldots, P_r$, et la permutation paramétrable $\widetilde{P}_0$ sont toutes uniformément aléatoires et indépendantes, de sorte que, d'après (2.2) et (2.3), la probabilité d'obtenir une quelconque transcription atteignable des requêtes $\tau = (\mathcal{Q}_C, \mathcal{Q}_{\mathbf{P}}, \mathbf{k})$ dans le monde réel vaut

$$\Pr[T_{\mathrm{id}} = \tau] = \frac{1}{|\mathcal{K}|} \times \left(\frac{1}{(N)_{q_p}}\right)^r \times \prod_{i=1}^{m} \frac{1}{(N)_{q_i}}.$$

Dans le monde réel, la probabilité d'obtenir τ est

$$\Pr[T_{\mathrm{re}} = \tau] = \frac{1}{|\mathcal{K}|} \times \left(\frac{1}{(N)_{q_p}}\right)^r \times \Pr\left[\mathbf{P} \leftarrow_{\$} (\mathsf{P}(n))^r : \mathsf{TEM}_{\mathbf{k}}^{\mathbf{P}} \vdash \mathcal{Q}_C \,\middle|\, \mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}}\right].$$

Soit

$$\mathsf{p}(\tau) \stackrel{\mathrm{def}}{=} \Pr\left[\mathbf{P} \leftarrow_{\$} (\mathsf{P}(n))^r : \mathsf{TEM}_{\mathbf{k}}^{\mathbf{P}} \vdash \mathcal{Q}_C \,\middle|\, \mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}}\right].$$

Alors on a

$$\frac{\Pr[T_{\mathrm{re}} = \tau]}{\Pr[T_{\mathrm{id}} = \tau]} = \mathsf{p}(\tau) \Big/ \prod_{i=1}^{m} \frac{1}{(N)_{q_i}}. \tag{2.4}$$

Par conséquent, l'application du lemme 1 requiert trois étapes : tout d'abord, définir les bonnes et les mauvaises transcriptions, puis majorer la probabilité d'obtenir une mauvaise transcription dans le monde idéal, et finalement minorer la probabilité $\mathsf{p}(\tau)$ dans le monde réel lorsque τ est une bonne transcription afin d'utiliser l'équation (2.4).

Chapitre 3

Un mixage non-linéaire de la clé et du *tweak*

3.1 Présentation de la construction

Ce chapitre est consacré à l'étude d'une variante paramétrable du schéma d'Even-Mansour inspirée de la construction de Liskov, Rivest et Wagner [LRW02]. On rappelle que cette construction générique permet, à partir d'une famille ε-AXU de fonctions de $\mathcal{T}$ vers $\{0,1\}^n$, notée $\mathcal{H}$, et d'un algorithme de chiffrement par blocs E travaillant sur des blocs de n bits et dont l'ensemble de clés est noté $\mathcal{K}$, de construire un algorithme de chiffrement par blocs paramétrable $\widetilde{E}$, sûr jusqu'à la borne des anniversaires, ayant pour ensemble de clés le produit cartésien $\mathcal{H} \times \mathcal{K}$, pour ensemble de *tweaks* $\mathcal{T}$ et pour ensemble de messages $\{0,1\}^n$ tel que, pour tout quadruplet $(h_1, k_1, t, x) \in \mathcal{H} \times \mathcal{K} \times \mathcal{T} \times \{0,1\}^n$,

$$\widetilde{E}_{(h_1,k_1)}(t,x) = h_1(t) \oplus E_{k_1}(h_1(t) \oplus x).$$

Cette construction peut être itérée plusieurs fois, en utilisant deux nouvelles clés indépendantes pour chaque tour, afin d'augmenter la sécurité de l'algorithme final, qui dépend également du choix de l'algorithme E. En utilisant les résultats de sécurité existants, pour obtenir une sécurité convenable, il serait nécessaire de remplacer E par un schéma d'Even-Mansour à plusieurs tours, aboutissant à un algorithme ayant de mauvaises performances. Il paraît plus prometteur de combiner la construction LRW d'une part et le schéma d'Even-Mansour à un tour d'autre part. On obtient alors un schéma de chiffrement par blocs paramétrable E sûr jusqu'à la borne des anniversaires, ayant pour ensemble de clés le produit cartésien $\mathcal{H} \times \{0,1\}^n \times \{0,1\}^n$, pour ensemble de *tweaks* $\mathcal{T}$ et pour ensemble de messages $\{0,1\}^n$ tel que, pour tout quadruplet $(h_1, k_1, k_2, t, x) \in \mathcal{H} \times \{0,1\}^n \times \{0,1\}^n \times \mathcal{T} \times \{0,1\}^n$, défini comme suit :

$$E_{(h_1,k_1)}(t,x) = h_1(t) \oplus k_2 \oplus P(h_1(t) \oplus k_1 \oplus x).$$

Cette construction n'est cependant pas optimale : si l'on demande à la famille de fonctions $\mathcal{H}$ d'être uniforme, nous allons voir que les clés k_1 et k_2 peuvent être supprimées.

Définition 6. *Une famille $\mathcal{F}$ de fonctions de A vers B est dite uniforme lorsque, pour tout élément a dans A et tout élément b dans B, on a*

$$\Pr\left[f \leftarrow_{\$} \mathcal{F} \,:\, f(a) = b\right] = \frac{1}{|B|}.$$

Plus précisément, fixons pour le reste de ce chapitre un entier naturel n et une famille uniforme et ε-AXU de fonctions d'un ensemble non-vide $\mathcal{T}$ vers $\{0,1\}^n$ notée $\mathcal{H}$. Notons également $N = 2^n$. La construction que nous allons étudier est le schéma d'Even-Mansour paramétrable $\mathsf{TEM}[n, r, \mathbf{f}_{\mathcal{H}}]$, pour lequel la famille $\mathbf{f}_{\mathcal{H}} = (f_0, \ldots, f_r)$ de fonctions de dérivation de clés est définie sur l'ensemble $\mathcal{H} \times \mathcal{T}$ par, pour $i = 0, \ldots, r$,

$$\forall\, ((h_1, \ldots, h_r), t) \in \mathcal{H}^r \times \mathcal{T},\ f_i\left((h_1, \ldots, h_r), t\right) = h_{i+1}(t) \oplus h_i(t),$$

où, par convention, $h_0(t) = h_{r+1}(t) = 0$ pour tout $t \in \mathcal{T}$. Nous allons prouver que cette construction dispose de fortes garanties de sécurité : un tour est suffisant pour atteindre le niveau de la borne des anniversaires, tandis que deux tours suffisent pour que la construction soit sécurisée tant que le nombre de requêtes de l'adversaire est petit devant $2^{2n/3}$, où n est la taille du bloc. Nous prouvons finalement une borne asymptotique qui tend vers la sécurité optimale à mesure que le nombre r de tours croît. Ces travaux ont fait l'objet d'une publication à la conférence CRYPTO 2015 [CLS15].

3.2 Preuve de sécurité pour 1 tour

Nous considérons ici la construction $\mathsf{TEM}[n, 1, \mathbf{f}_{\mathcal{H}}]$: pour tout *tweak* $t \in \mathcal{T}$ et toute clé $h_1 \in \mathcal{H}$, le chiffré d'un message $m \in \{0,1\}^n$ quelconque est donné par :

$$\mathsf{TEM}_{h_1}^{P_1}(t, x) = h_1(t) \oplus P_1(h_1(t) \oplus x)$$

où P_1 est une permutation aléatoire publique. La sécurité de cette construction est donnée par le théorème suivant.

Théorème 4. *Soient q_c et q_p deux entiers naturels. On a*

$$\mathbf{Adv}^{\mathrm{cca}}_{\mathsf{TEM}[n,1,\mathbf{f}_{\mathcal{H}}]}(q_c, q_p) \leq q_c^2 \varepsilon + \frac{2 q_c q_p}{N}.$$

La preuve que nous allons donner repose sur la technique des coefficients H présentée dans la section 1.4. Elle peut être vue comme un bon échauffement pour les preuves suivantes. Définissons tout d'abord les mauvaises transcriptions.

Définition 7. *Une transcription atteignable $\tau = (\mathcal{Q}_C, \mathcal{Q}_{P_1}, h_1)$ est dite mauvaise si une des quatre conditions suivantes est satisfaite :*

(C-1) il existe deux requêtes distinctes $(t, x, y), (t', x', y') \in \mathcal{Q}_C$ telles que $h_1(t) \oplus h_1(t') = x \oplus x'$;

(C-2) il existe deux requêtes distinctes $(t, x, y), (t', x', y') \in \mathcal{Q}_C$ *telles que* $h_1(t) \oplus h_1(t') = y \oplus y'$ *;*

(C-3) il existe $(t, x, y) \in \mathcal{Q}_C$ *et* $(u, v) \in \mathcal{Q}_{P_1}$ *tels que* $x \oplus h_1(t) = u$ *;*

(C-4) il existe $(t, x, y) \in \mathcal{Q}_C$ *et* $(u, v) \in \mathcal{Q}_{P_1}$ *tels que* $y \oplus h_1(t) = v$.

Dans le cas contraire τ *est une bonne transcription. On note* Θ_{good}*, respectivement* Θ_{bad} *l'ensemble des bonnes, respectivement des mauvaises transcriptions.* ◇

Autrement dit, une transcription est bonne lorsqu'il n'y a aucune collision simple entre des requêtes, aussi bien en entrée qu'en sortie de la permutation. On remarquera qu'une telle définition des mauvaises transcriptions n'est possible que grâce au fait que nous révélons la clé h_1 utilisée à l'adversaire une fois ses requêtes terminées.

Commençons par majorer la probabilité d'obtenir une mauvaise transcription dans le monde idéal.

Lemme 13.

$$\Pr[T_{\text{id}} \in \Theta_{\text{bad}}] \leq q_c^2 \varepsilon + \frac{2 q_c q_p}{N}.$$

Démonstration. Soit $(\mathcal{Q}_C, \mathcal{Q}_{P_1})$ une transcription des requêtes atteignables. Notons que, dans le monde idéal, la clé h_1 est tirée uniformément aléatoirement dans $\mathcal{H}$, indépendamment de la transcription des requêtes. Fixons une paire de requêtes deux à deux distinctes (t, x, y) et $(t', x', y') \in \mathcal{Q}_C$. Puisque $\mathcal{H}$ est une famille de fonctions ayant la propriété ε-AXU, nous avons

$$\Pr[h_1 \leftarrow_{\$} \mathcal{H} : h_1(t) \oplus h_1(t') = x \oplus x' \vee h_1(t) \oplus h_1(t') = y \oplus y'] \leq 2\varepsilon.$$

On note que cette inégalité reste vraie si $t = t'$. En effet, dans ce cas, nécessairement $x \neq x'$ et $y \neq y'$ car nous avons supposé que $\mathcal{D}$ n'effectue jamais de requêtes redondantes. Ainsi, en sommant sur les $q_c(q_c - 1)/2$ paires de requêtes possibles, la probabilité que les conditions (C-1) ou (C-2) soient satisfaites est au plus $q_c^2 \varepsilon$.

De plus, pour tout $(t, x, y) \in \mathcal{Q}_C$ et tout $(u, v) \in \mathcal{Q}_{P_1}$, la probabilité, quand h_1 est tirée aléatoirement, que $h_1(t) = x \oplus u$ ou que $h_1(t) = y \oplus v$ est inférieure à $2/N$ puisque $\mathcal{H}$ est une famille uniforme de fonctions. Ainsi, la probabilité que les conditions (C-3) ou (C-4) soient satisfaites est inférieure à $2q_c q_p/N$. □

Analysons ensuite les bonnes transcriptions.

Lemme 14. *Pour toute bonne transcription* τ*, on a*

$$\frac{\Pr[T_{\text{re}} = \tau]}{\Pr[T_{\text{id}} = \tau]} \geq 1.$$

Démonstration. Soit τ une bonne transcription. D'après (2.4), nous devons minorer la probabilité suivante :

$$\mathsf{p}(\tau) \stackrel{\text{def}}{=} \Pr[P_1 \leftarrow_{\$} \mathsf{P}(n) : \forall (t, x, y) \in \mathcal{Q}_C, P_1(x \oplus h_1(t)) = y \oplus h_1(t) \mid P_1 \vdash \mathcal{Q}_{P_1}].$$

Comme τ est une bonne transcription, toutes les valeurs $x \oplus h_1(t)$ lorsque (t, x, y) parcourt $\mathcal{Q}_C$ sont deux à deux distinctes car sinon τ vérifierait la condition (C-1),

et également deux à deux distinctes des valeurs u pour $(u, v) \in \mathcal{Q}_{P_1}$ car sinon τ vérifierait la condition (C-3). De même, toutes les valeurs $y \oplus h_1(t)$ pour $(t, x, y) \in \mathcal{Q}_C$ sont deux à deux distinctes car sinon τ vérifierait la condition (C-2), et également distinctes des valeurs v pour $(u, v) \in \mathcal{Q}_{P_1}$ car sinon τ vérifierait la condition (C-4). Ceci entraîne clairement que

$$\mathsf{p}(\tau) = \frac{1}{(N - q_p)_{q_c}},$$

d'où, d'après (2.4), on a

$$\frac{\Pr[T_{\mathrm{re}} = \tau]}{\Pr[T_{\mathrm{id}} = \tau]} = \frac{\prod_{i=1}^{m} (N)_{q_i}}{(N - q_p)_{q_c}} \geq \frac{(N)_{q_c}}{(N - q_p)_{q_c}} \geq 1.$$

□

La preuve du théorème 4 s'obtient en combinant les lemmes 1 (en choisissant $\varepsilon_1 = 0$), 13, et 14.

3.3 Preuve de sécurité pour 2 tours

3.3.1 Résultat

Nous considérons ici la construction $\mathsf{TEM}[n, 2, \mathbf{f}_{\mathcal{H}}]$: pour tout *tweak* $t \in \mathcal{T}$ et toute clé $(h_1, h_2) \in \mathcal{H}^2$, le chiffré d'un message $m \in \{0,1\}^n$ quelconque est donné par :

$$\mathsf{TEM}^{P_1,P_2}_{(h_1,h_2)}(t, x) = h_2(t) \oplus P_2\Big(h_2(t) \oplus h_1(t) \oplus P_1(h_1(t) \oplus x)\Big)$$

où P_1, P_2 sont deux permutations aléatoires publiques et indépendantes. La sécurité de cette construction est donnée par le théorème suivant.

Théorème 5. *Soient q_p et q_c deux entiers naturels. On suppose que $q_p + 3q_c \leq N/2$ et $q_c \leq \min\{N^{2/3}, \varepsilon^{-2/3}\}$. Alors*

$$\mathbf{Adv}^{\mathrm{cca}}_{\mathsf{TEM}[n,2,\mathcal{H}]}(q_c, q_p) \leq \frac{29\sqrt{q_c}q_p}{N} + \varepsilon\sqrt{q_c}q_p + 6\varepsilon q_c^{3/2} + \frac{30q_c^{3/2}}{N}.$$

En particulier, si on suppose que $\mathcal{H}$ est XOR universelle (i.e., $\varepsilon = 2^{-n}$), on peut voir que la construction TEM à deux tours assure un bon niveau de sécurité jusqu'à approximativement $2^{2n/3}$ requêtes de l'adversaire. En fait, la construction TEM à deux tours reste sûre tant que le nombre q_c de requêtes à la construction est petit devant $2^{2n/3}$ et le nombre de requêtes à la permutation est petit devant $2^n/\sqrt{q_c}$.

La preuve repose toujours sur la technique des coefficients H. Comme d'habitude, nous commencerons par définir les mauvaises transcriptions et majorer la probabilité qu'elles apparaissent dans le monde idéal dans la section 3.3.2, puis nous montrerons dans la section 3.3.3 que, pour chaque bonne transcription, les probabilités d'obtenir cette transcription dans le monde réel et dans le monde idéal sont suffisamment proches.

3.3.2 Description des mauvaises transcriptions

Soit $\tau = (\mathcal{Q}_C, \mathcal{Q}_{P_1}, \mathcal{Q}_{P_2}, (h_1, h_2))$ une transcription atteignable. On note $|\mathcal{Q}_C| = q_c$ et $|\mathcal{Q}_{P_1}| = |\mathcal{Q}_{P_2}| = q_p$. Soient

$$U_1 = \{u_1 \in \{0,1\}^n : (u_1, v_1) \in \mathcal{Q}_{P_1}\}, \qquad V_1 = \{v_1 \in \{0,1\}^n : (u_1, v_1) \in \mathcal{Q}_{P_1}\},$$
$$U_2 = \{u_2 \in \{0,1\}^n : (u_2, v_2) \in \mathcal{Q}_{P_2}\}, \qquad V_2 = \{v_2 \in \{0,1\}^n : (u_2, v_2) \in \mathcal{Q}_{P_2}\},$$

les ensembles de définition et les images de $\mathcal{Q}_{P_1}$ et $\mathcal{Q}_{P_2}$ respectivement. Pour tout u et $v \in \{0,1\}^n$, soient

$$X_u = \{(t, x, y) \in \mathcal{Q}_C : x \oplus h_1(t) = u\},$$
$$Y_v = \{(t, x, y) \in \mathcal{Q}_C : y \oplus h_2(t) = v\}.$$

Définissons quatre quantités qui caractérisent le nombre de collisions entre des requêtes en entrée de P_1 et en sortie de P_2 forcées par la transcription τ :

$$\alpha_1 \stackrel{\text{def}}{=} |\{(t, x, y) \in \mathcal{Q}_C : x \oplus h_1(t) \in U_1\}|,$$
$$\alpha_2 \stackrel{\text{def}}{=} |\{(t, x, y) \in \mathcal{Q}_C : y \oplus h_2(t) \in V_2\}|,$$
$$\beta_1 \stackrel{\text{def}}{=} |\{(t, x, y) \in \mathcal{Q}_C : \exists (t', x', y') \neq (t, x, y), x \oplus h_1(t) = x' \oplus h_1(t')\}|,$$
$$\beta_2 \stackrel{\text{def}}{=} |\{(t, x, y) \in \mathcal{Q}_C : \exists (t', x', y') \neq (t, x, y), y \oplus h_2(t) = y' \oplus h_2(t')\}|.$$

Intuitivement, α_1 (respectivement α_2) est le nombre de requête (t, x, y) à la construction qui collisionnent avec une requête (u_1, v_1) à la première permutation (respectivement avec une requête (u_2, v_2) à la seconde permutation), et β_1 (respectivement β_2) est le nombre de requêtes (t, x, y) à la construction qui collisionnent avec une autre requête (t', x', y') en entrée de P_1 (respectivement en sortie de P_2). Remarquons que l'on a également

$$\beta_1 = \sum_{\substack{u \in \{0,1\}^n : \\ |X_u| > 1}} |X_u|, \qquad \beta_2 = \sum_{\substack{v \in \{0,1\}^n : \\ |Y_v| > 1}} |Y_v|. \tag{3.1}$$

Définition 8. *On dit qu'une transcription atteignable τ est mauvaise si au moins l'une des conditions suivantes est satisfaite (voir Figure 3.1 pour un diagramme représentant les dix premières conditions) :*

(C-1) il existe $(t, x, y) \in \mathcal{Q}_C$, $u_1 \in U_1$, et $v_2 \in V_2$ tels que $x \oplus h_1(t) = u_1$ et $y \oplus h_2(t) = v_2$;

(C-2) il existe $(t, x, y) \in \mathcal{Q}_C$, $(u_1, v_1) \in \mathcal{Q}_{P_1}$, et $u_2 \in U_2$ tels que $x \oplus h_1(t) = u_1$ et $v_1 \oplus h_1(t) \oplus h_2(t) = u_2$;

(C-3) il existe $(t, x, y) \in \mathcal{Q}_C$, $(u_2, v_2) \in \mathcal{Q}_{P_2}$, et $v_1 \in V_1$ tels que $y \oplus h_2(t) = v_2$ et $v_1 \oplus h_1(t) \oplus h_2(t) = u_2$;

(C-4) il existe $(t, x, y), (t', x', y'), (t'', x'', y'') \in \mathcal{Q}_C$, tels que (t, x, y) est différent de (t', x', y') et de (t'', x'', y''), $x \oplus h_1(t) = x' \oplus h_1(t')$ et $y \oplus h_2(t) = y'' \oplus h_2(t'')$;

(C-5) il existe $(t, x, y) \neq (t', x', y') \in \mathcal{Q}_C$ tels que $x \oplus h_1(t) = x' \oplus h_1(t')$ et $h_1(t) \oplus h_2(t) = h_1(t') \oplus h_2(t')$;

(C-6) *il existe* $(t,x,y) \neq (t',x',y') \in \mathcal{Q}_C$ *tels que* $y \oplus h_2(t) = y' \oplus h_2(t')$ *et* $h_1(t) \oplus h_2(t) = h_1(t') \oplus h_2(t')$ *;*

(C-7) *il existe* $(t,x,y) \neq (t',x',y') \in \mathcal{Q}_C$ *et* $u_1 \in U_1$ *tels que* $y \oplus h_2(t) = y' \oplus h_2(t')$ *et* $x \oplus h_1(t) = u_1$ *;*

(C-8) *il existe* $(t,x,y) \neq (t',x',y') \in \mathcal{Q}_C$ *et* $v_2 \in V_2$ *tels que* $x \oplus h_1(t) = x' \oplus h_1(t')$ *et* $y \oplus h_2(t) = v_2$ *;*

(C-9) *il existe* $(t,x,y) \neq (t',x',y') \in \mathcal{Q}_C$, $(u_1,v_1),(u_1',v_1') \in \mathcal{Q}_{P_1}$ *tels que* $x \oplus h_1(t) = u_1$, $x' \oplus h_1(t') = u_1'$ *et* $v_1 \oplus h_1(t) \oplus h_2(t) = v_1' \oplus h_1(t') \oplus h_2(t')$ *;*

(C-10) *il existe* $(t,x,y) \neq (t',x',y') \in \mathcal{Q}_C$, $(u_2,v_2),(u_2',v_2') \in \mathcal{Q}_{P_2}$ *tels que* $y \oplus h_2(t) = v_2$, $y' \oplus h_2(t') = v_2'$ *et* $u_2 \oplus h_1(t) \oplus h_2(t) = u_2' \oplus h_1(t') \oplus h_2(t')$ *;*

(C-11) $\alpha_1 \geq \sqrt{q_c}$ *;*

(C-12) $\alpha_2 \geq \sqrt{q_c}$ *;*

(C-13) $\beta_1 \geq \sqrt{q_c}$ *;*

(C-14) $\beta_2 \geq \sqrt{q_c}$.

Dans le cas contraire, τ *est une bonne transcription. Notons* Θ_{good}*, respectivement* Θ_{bad}*, l'ensemble des bonnes, respectivement des mauvaises transcriptions.* $\diamond$

Commençons par majorer la probabilité d'obtenir une mauvaise transcription dans le monde idéal.

Lemme 15. *Pour toute paire d'entiers naturels* (q_c, q_p)*, on a*

$$\Pr[T_{\text{id}} \in \Theta_{\text{bad}}] \leq \frac{3q_cq_p^2}{N^2} + 2\varepsilon^2 q_c^3 + \frac{\varepsilon q_c^2 q_p}{N} + \frac{2\sqrt{q_c}q_p}{N} + 4\varepsilon q_c^{3/2}.$$

Démonstration. Soit $(\mathcal{Q}_C, \mathcal{Q}_{P_1}, \mathcal{Q}_{P_2})$ une transcription atteignable des requêtes. On rappelle que, dans le monde idéal, (h_1, h_2) est tirée aléatoirement, indépendamment de la transcription des requêtes. Nous majorons les probabilités que les quatorze conditions soient satisfaites à tour de rôle. Notons Θ_i l'ensemble des transcriptions atteignables satisfaisant la condition (C-i).

Conditions (C-1), (C-2), et (C-3). Considérons la condition (C-1). Pour toute requête $(t,x,y) \in \mathcal{Q}_C$, tout $u_1 \in U_1$, et tout $v_2 \in V_2$, on a, en utilisant l'uniformité de $\mathcal{H}$ et puisque h_1 et h_2 sont tirés indépendamment,

$$\Pr\left[\Big(h_1(t) = x \oplus u_1\Big) \wedge \Big(h_2(t) = y \oplus v_2\Big)\right] = \frac{1}{N^2}.$$

Ainsi, en sommant sur les $q_cq_p^2$ choix possibles pour (t,x,y), u_1, et v_1, on obtient

$$\Pr[T_{\text{id}} \in \Theta_1] \leq \frac{q_cq_p^2}{N^2}.$$

De même, pour les conditions (C-2) et (C-3), on a

$$\Pr\left[T_{\text{id}} \in \Theta_2\right] \leq \frac{q_cq_p^2}{N^2}$$

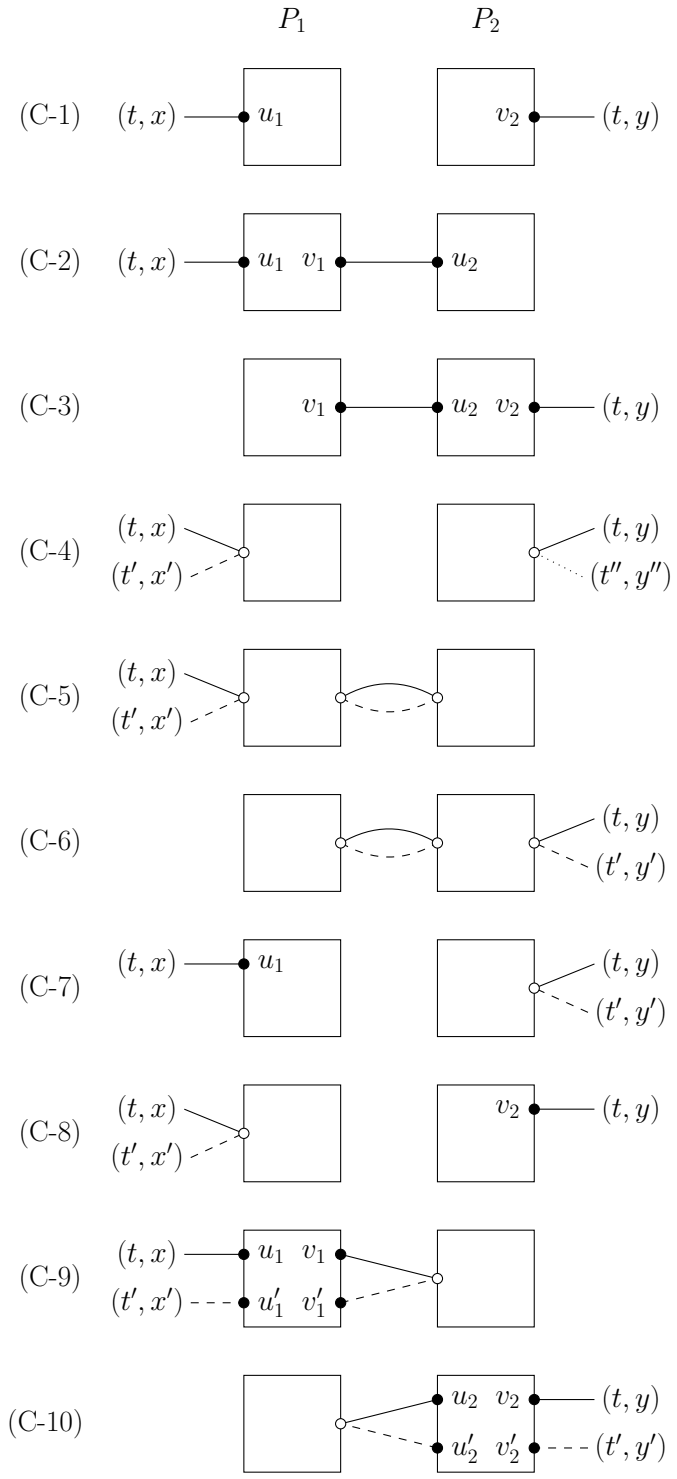

FIGURE 3.1 – Les dix conditions caractérisant les collisions qui définissent une mauvaise transcription. Les points noirs correspondent aux paires $(u_1, v_1) \in \mathcal{Q}_{P_1}$ ou $(u_2, v_2) \in \mathcal{Q}_{P_2}$. Notons que, pour (C-4), il est possible que $(t', x') = (t'', x'')$, pour (C-9) (respectivement (C-10)) on peut avoir l'égalité $(u_1, v_1) = (u'_1, v'_1)$ (respectivement $(u_2, v_2) = (u'_2, v'_2)$).

$$\Pr\left[T_{\mathrm{id}} \in \Theta_3\right] \leq \frac{q_c q_p^2}{N^2}.$$

Condition (C-4). Pour tout triplet $((t,x,y),(t',x',y'),(t'',x'',y'')) \in \mathcal{Q}_C^3$ où (t,x,y) est différent de (t',x',y') et de (t'',x'',y''), on a, puisque $\mathcal{H}$ est ε-AXU et les clés h_1 et h_2 sont tirées indépendamment,

$$\Pr\left[\Big(h_1(t) \oplus h_1(t') = x \oplus x'\Big) \wedge \Big(h_2(t) \oplus h_2(t'') = y \oplus y''\Big)\right] \leq \varepsilon^2.$$

Notons que cette inégalité reste vraie lorsque $t = t'$ (respectivement $t = t''$) car, dans ce cas, nécessairement $x \neq x'$ (respectivement $y \neq y''$) d'après l'hypothèse que le distingueur ne fait jamais de requête inutile. Ainsi, en sommant sur les (au plus) q_c^3 choix possibles pour (t,x,y), (t',x',y'), (t'',x'',y''), on obtient

$$\Pr\left[T_{\mathrm{id}} \in \Theta_4\right] \leq \varepsilon^2 q_c^3.$$

Conditions (C-5) et (C-6). Pour toutes requêtes distinctes $(t,x,y) \neq (t',x',y') \in \mathcal{Q}_C$, on a, par la propriété ε-AXU de $\mathcal{H}$ et puisque h_1 et h_2 sont tirées indépendamment

$$\Pr\left[\Big(h_1(t) \oplus h_1(t') = x \oplus x'\Big) \wedge \Big(h_2(t) \oplus h_2(t') = h_1(t) \oplus h_1(t')\Big)\right] \leq \varepsilon^2.$$

Ainsi, en sommant sur les $q_c(q_c-1)/2$ paires de requêtes distinctes possibles, on obtient

$$\Pr\left[T_{\mathrm{id}} \in \Theta_5\right] \leq \frac{\varepsilon^2 q_c^2}{2}.$$

De même,

$$\Pr\left[T_{\mathrm{id}} \in \Theta_6\right] \leq \frac{\varepsilon^2 q_c^2}{2}.$$

Conditions (C-7) et (C-8). Pour toutes requêtes $(t,x,y) \neq (t',x',y') \in \mathcal{Q}_C$ et tout $u_1 \in U_1$, on a, en utilisant la propriété ε-AXU et l'uniformité de $\mathcal{H}$ et puisque h_1 et h_2 sont tirées indépendamment,

$$\Pr\left[\Big(h_2(t) \oplus h_2(t') = y \oplus y'\Big) \wedge \Big(h_1(t) = x \oplus u_1\Big)\right] \leq \frac{\varepsilon}{N}.$$

Puis, en sommant sur les choix possibles de $(t,x,y) \neq (t',x',y')$ et u_1,

$$\Pr[T_{\mathrm{id}} \in \Theta_7] \leq \frac{\varepsilon q_c^2 q_p}{2N}.$$

De même,

$$\Pr[T_{\mathrm{id}} \in \Theta_8] \leq \frac{\varepsilon q_c^2 q_p}{2N}.$$

Conditions (C-9), (C-10), (C-11), et (C-12). Nous allons étudier les conditions (C-9) et (C-11) simultanément, en utilisant le fait que

$$\Pr\left[T_{\mathrm{id}} \in \Theta_9 \cup \Theta_{11}\right] = \Pr\left[T_{\mathrm{id}} \in \Theta_{11}\right] + \Pr\left[T_{\mathrm{id}} \in \Theta_9 \setminus \Theta_{11}\right].$$

Pour majorer $\Pr[T_{\text{id}} \in \Theta_{11}]$, on considère α_1 comme une variable aléatoire sur le choix aléatoire de h_1 (puisque α_1 ne dépend pas de h_2). Tout d'abord, à cause de l'uniformité de $\mathcal{H}$,

$$\mathbb{E}[\alpha_1] = \sum_{(t,x,y)\in\mathcal{Q}_C} \sum_{u_1\in U_1} \Pr[x \oplus h_1(t) = u_1] = \frac{q_c q_p}{N},$$

d'où, par l'inégalité de Markov,

$$\Pr[T_{\text{id}} \in \Theta_{11}] \leq \frac{\sqrt{q_c} q_p}{N}.$$

Fixons $h_1' \in \mathcal{H}$ tel que, quand $h_1 = h_1'$, $\alpha_1 < \sqrt{q_c}$, et fixons des requêtes $(t,x,y) \neq (t',x',y') \in \mathcal{Q}_C$, $(u_1,v_1),(u_1',v_1') \in \mathcal{Q}_{P_1}$ telle que $x \oplus h_1(t) = u_1$ et $x' \oplus h_1(t') = u_1'$. On note que puisque $\alpha_1 < \sqrt{q_c}$, il y a au plus $\frac{q_c}{2}$ tels uplets de requêtes. Alors

$$\Pr\left[\left(h_1 = h_1'\right) \wedge \left(h_2(t) \oplus h_2(t') = v_1 \oplus h_1(t) \oplus v_1' \oplus h_1(t')\right)\right] \leq \frac{\varepsilon}{|\mathcal{H}|},$$

et, en sommant sur tous les h_1 tels que $\alpha_1 < \sqrt{q_c}$ et tous les uplets de requêtes de la forme définie précédemment, on a

$$\Pr[T_{\text{id}} \in \Theta_9 \setminus \Theta_{11}] \leq \frac{\varepsilon q_c}{2}.$$

Enfin,

$$\Pr[T_{\text{id}} \in \Theta_9 \cup \Theta_{11}] \leq \frac{\sqrt{q_c} q_p}{N} + \frac{\varepsilon q_c}{2}.$$

De même,

$$\Pr[T_{\text{id}} \in \Theta_{10} \cup \Theta_{12}] \leq \frac{\sqrt{q_c} q_p}{N} + \frac{\varepsilon q_c}{2}.$$

Conditions (C-13) et (C-14). Pour tout $u \in \{0,1\}^n$, on voit $|X_u|$ comme une variable aléatoire sur le choix aléatoire de h_1. On définit également la variable aléatoire suivante :

$$C = |\{((t,x,y),(t',x',y')) \in \mathcal{Q}_C^2, (t,x,y) \neq (t',x',y') : x \oplus h_1(t) = x' \oplus h_1(t')\}|.$$

Alors, par définition de β_1,

$$\beta_1 = |\{(t,x,y) \in \mathcal{Q}_C : \exists (t',x',y') \neq (t,x,y), x \oplus h_1(t) = x' \oplus h_1(t')\}| \leq C.$$

Ainsi, $\Pr[T_{\text{id}} \in \Theta_{13}] \leq \Pr\left[C \geq \sqrt{q_c}\right]$. Remarquons que

$$\mathbb{E}[C] = \sum_{(t,x,y)\neq(t',x',y')} \Pr[x \oplus h_1(t) = x' \oplus h_1(t')] \leq \varepsilon q_c^2.$$

D'après l'inégalité de Markov,

$$\Pr[T_{\text{id}} \in \Theta_{13}] \leq \varepsilon q_c^{3/2}.$$

De même,

$$\Pr[T_{\text{id}} \in \Theta_{14}] \leq \varepsilon q_c^{3/2}.$$

□

3.3.3 Étude des bonnes transcriptions

Il faut maintenant étudier les bonnes transcriptions. Au cours des calculs, nous aurons besoin du lemme technique suivant, adapté de [CLL+14].

Lemme 16. *Soient N, a, b, c trois entiers naturels tels que $a+b \leq N/2$ et $a+c \leq N/2$. On a*

$$\frac{(N)_a(N-b-c)_a}{(N-b)_a(N-c)_a} \geq 1 - \frac{4abc}{N^2}.$$

Démonstration. On a

$$\begin{aligned}
\frac{(N)_a(N-b-c)_a}{(N-b)_a(N-c)_a} &= \prod_{i=0}^{a-1} \frac{(N-i)(N-b-c-i)}{(N-b-i)(N-c-i)} \\
&= \prod_{i=0}^{a-1} \frac{N^2 - N(b+c+2i) + i(b+c+i)}{N^2 - N(b+c+2i) + i(b+c+i) + bc} \\
&= \prod_{i=0}^{a-1} \left(1 - \frac{bc}{N^2 - N(b+c+2i) + i(b+c+i) + bc}\right) \\
&= \prod_{i=0}^{a-1} \left(1 - \frac{bc}{(N-b-i)(N-c-i)}\right) \\
&\geq \prod_{i=0}^{a-1} \left(1 - \frac{bc}{(N-b-a)(N-c-a)}\right) \\
&\geq 1 - \frac{abc}{(N-b-a)(N-c-a)} \\
&\geq 1 - \frac{4abc}{N^2},
\end{aligned}$$

car, par hypothèse, $a + b \leq N/2$ et $a + c \leq N/2$. □

Nous pouvons alors entamer l'étude des bonnes transcriptions.

Lemme 17. *Soient q_c et q_p deux entiers naturels tels que $q_p + 3q_c \leq N/2$. Pour toute bonne transcription τ, on a*

$$\frac{\Pr\left[T_{\mathrm{re}} = \tau\right]}{\Pr\left[T_{\mathrm{id}} = \tau\right]} \geq 1 - \left(\frac{4q_c(q_p + 2q_c)^2}{N^2} + \frac{14q_c^{3/2} + 4\sqrt{q_c}q_p}{N}\right).$$

Démonstration. Soit $\tau = (\mathcal{Q}_C, \mathcal{Q}_{P_1}, \mathcal{Q}_{P_2}, (h_1, h_2))$ une bonne transcription. D'après l'équation (2.4), nous devons minorer

$$\mathsf{p}(\tau) \stackrel{\text{def}}{=} \Pr\left[P_1, P_2 \leftarrow_{\$} \mathsf{P}(n) : \mathsf{TEM}^{P_1,P_2}_{(h_1,h_2)} \vdash \mathcal{Q}_C \middle| P_1 \vdash \mathcal{Q}_{P_1} \wedge P_2 \vdash \mathcal{Q}_{P_2}\right].$$

Notation. Nous allons grouper les requêtes à la construction selon le type de collision dans lequel ils sont impliqués. Concrètement, on définit (voir aussi la Figure 3.2 pour un diagramme de ces ensembles de requêtes)

$$\mathcal{Q}_{U_1} = \{(t, x, y) \in \mathcal{Q}_C : x \oplus h_1(t) \in U_1\},$$

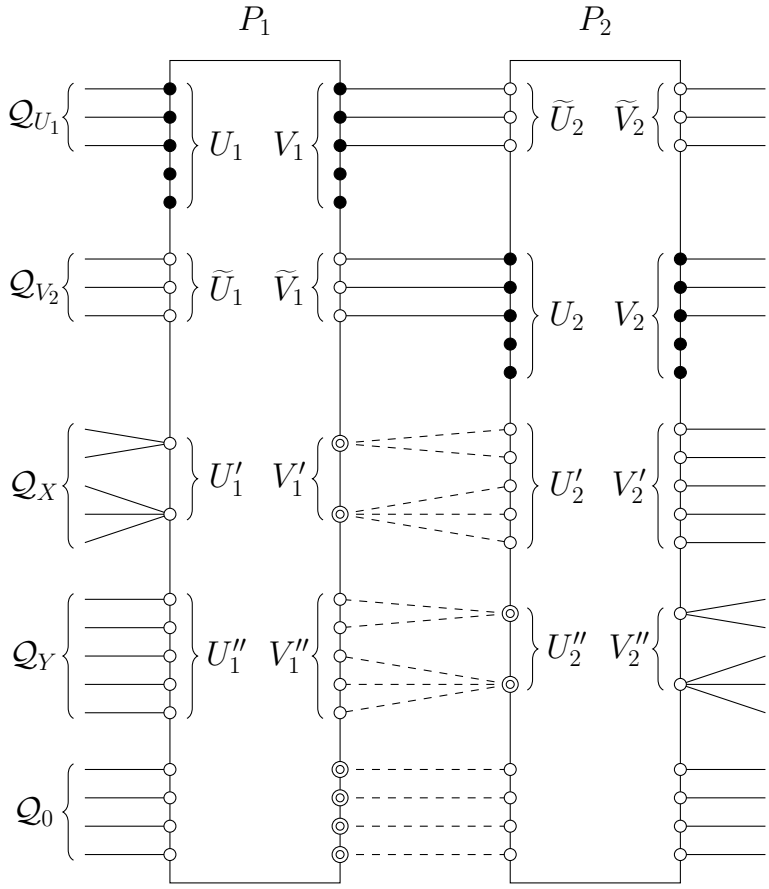

FIGURE 3.2 – Partition de $\mathcal{Q}_C$. Les points noirs correspondent aux valeurs fixées par les transcriptions des permutations internes $\mathcal{Q}_{P_1}$ et $\mathcal{Q}_{P_2}$. Les points blancs entourés correspondent aux valeurs $(v'_{1,i})_{1\le i\le\alpha'_1}$, $(u''_{2,i})_{1\le i\le\alpha''_2}$, et $(\mathbf{v}_{1,i,j})_{1\le i\le m,1\le j\le q'_i}$ sur lesquelles nous sommons pour la minoration de $\mathsf{p}''(\tau)$ dans la preuve du lemme 17.

$$\begin{aligned}
\mathcal{Q}_{V_2} &= \{(t,x,y)\in\mathcal{Q}_C : y\oplus h_2(t)\in V_2\},\\
\mathcal{Q}_X &= \{(t,x,y)\in\mathcal{Q}_C : |X_{x\oplus h_1(t)}|>1 \text{ et } x\oplus h_1(t)\notin U_1\},\\
\mathcal{Q}_Y &= \{(t,x,y)\in\mathcal{Q}_C : |Y_{y\oplus h_2(t)}|>1 \text{ et } y\oplus h_2(t)\notin V_2\},\\
\mathcal{Q}_0 &= \{(t,x,y)\in\mathcal{Q}_C : |X_{x\oplus h_1(t)}|=|Y_{y\oplus h_2(t)}|=1,\ x\oplus h_1(t)\notin U_1,\ \text{et } y\oplus h_2(t)\notin V_2\}.
\end{aligned}$$

Notons que, par définition, on a $|\mathcal{Q}_{U_1}|=\alpha_1$ et $|\mathcal{Q}_{V_2}|=\alpha_2$. Remarquons également que ces ensembles forment une partition de $\mathcal{Q}_C$:

- $\mathcal{Q}_{U_1}\cap\mathcal{Q}_{V_2}=\emptyset$ car sinon τ vérifierait (C-1),
- $\mathcal{Q}_{U_1}\cap\mathcal{Q}_Y=\emptyset$ car sinon τ vérifierait (C-7),
- $\mathcal{Q}_{V_2}\cap\mathcal{Q}_X=\emptyset$ car sinon τ vérifierait (C-8),
- $\mathcal{Q}_X\cap\mathcal{Q}_Y=\emptyset$ car sinon τ vérifierait (C-4),
- $\mathcal{Q}_{U_1}\cap\mathcal{Q}_X=\mathcal{Q}_{U_1}\cap\mathcal{Q}_0=\mathcal{Q}_{V_2}\cap\mathcal{Q}_Y=\mathcal{Q}_{V_2}\cap\mathcal{Q}_0=\mathcal{Q}_X\cap\mathcal{Q}_0=\mathcal{Q}_Y\cap\mathcal{Q}_0=\emptyset$ par définition.

On note respectivement E_{U_1}, E_{V_2}, E_X, E_Y, et E_0 les évènements $\mathsf{TEM}^{P_1,P_2}_{(h_1,h_2)}\vdash\mathcal{Q}_{U_1}$, $\mathcal{Q}_{V_2}$, $\mathcal{Q}_X$, $\mathcal{Q}_Y$, et $\mathcal{Q}_0$. Comme l'évènement $\mathsf{TEM}^{P_1,P_2}_{(h_1,h_2)}\vdash\mathcal{Q}_C$ est équivalent à

$\mathsf{E}_{U_1} \wedge \mathsf{E}_{V_2} \wedge \mathsf{E}_X \wedge \mathsf{E}_Y \wedge \mathsf{E}_0$, on a

$$\begin{aligned}\mathsf{p}(\tau) &= \Pr\left[\mathsf{TEM}^{P_1,P_2}_{(h_1,h_2)} \vdash \mathcal{Q}_C \middle| P_i \vdash \mathcal{Q}_{P_i}, i = 1,2\right] \\ &= \Pr\left[\mathsf{E}_{U_1} \wedge \mathsf{E}_{V_2} \wedge \mathsf{E}_X \wedge \mathsf{E}_Y \wedge \mathsf{E}_0 | P_i \vdash \mathcal{Q}_{P_i}, i = 1,2\right] \\ &= \mathsf{p}'(\tau) \cdot \mathsf{p}''(\tau), \end{aligned} \tag{3.2}$$

où

$$\begin{aligned}\mathsf{p}'(\tau) &= \Pr\left[\mathsf{E}_{U_1} \wedge \mathsf{E}_{V_2} | P_i \vdash \mathcal{Q}_{P_i}, i = 1,2\right] \\ \mathsf{p}''(\tau) &= \Pr\left[\mathsf{E}_X \wedge \mathsf{E}_Y \wedge \mathsf{E}_0 | \mathsf{E}_{U_1} \wedge \mathsf{E}_{V_2} \wedge (P_i \vdash \mathcal{Q}_{P_i}, i = 1,2)\right]\end{aligned}$$

et les probabilités sont prises sur le choix aléatoire de P_1 et P_2. Minorons maintenant $\mathsf{p}'(\tau)$ et $\mathsf{p}''(\tau)$ successivement.

Minoration de $\mathsf{p}'(\tau)$. Lorsque l'on conditionne par les évènements $(P_i \vdash \mathcal{Q}_{P_i}, i = 1,2)$, les permutations P_1 et P_2 sont fixées sur exactement q_p valeurs chacune. Pour tout $(t,x,y) \in \mathcal{Q}_{U_1}$, il existe un unique $(u_1, v_1) \in \mathcal{Q}_{P_1}$ tel que $x \oplus h_1(t) = u_1$, donc $P_1(x \oplus h_1(t))$ est bien défini (et égal à v_1). Dans la suite, on note (voir la Figure 3.2)

$$\begin{aligned}\widetilde{U}_2 &= \{P_1(x \oplus h_1(t)) \oplus h_1(t) \oplus h_2(t) : (t,x,y) \in \mathcal{Q}_{U_1}\} \\ \widetilde{V}_2 &= \{y \oplus h_2(t) : (t,x,y) \in \mathcal{Q}_{U_1}\}.\end{aligned}$$

Remarquons que toutes les valeurs définissant $\widetilde{U}_2$ sont deux à deux différentes, car sinon τ vérifierait (C-9). De même toutes les valeurs définissant $\widetilde{V}_2$ sont deux à deux distinctes car sinon τ vérifierait (C-7). De plus, constatons que U_2 et $\widetilde{U}_2$ sont disjoints car sinon τ vérifierait (C-2) ; V_2 et $\widetilde{V}_2$ sont également disjoints car sinon τ vérifierait (C-1). Ainsi, l'évènement E_{U_1} est équivalent à α_1 «nouvelles» équations distinctes sur P_2, ainsi

$$\Pr\left[\mathsf{E}_{U_1} | P_2 \vdash \mathcal{Q}_{P_2}\right] = \frac{1}{(N - q_p)_{\alpha_1}}. \tag{3.3}$$

De même, pour chaque $(t,x,y) \in \mathcal{Q}_{V_2}$, il existe un unique $(u_2, v_2) \in \mathcal{Q}_{P_2}$ tel que $y \oplus h_2(t) = v_2$, donc $P_2^{-1}(y \oplus h_2(t))$ est bien défini et égal à u_2. Dans la suite, on note (voir la Figure 3.2)

$$\begin{aligned}\widetilde{V}_1 &= \{P_2^{-1}(y \oplus h_2(t)) \oplus h_1(t) \oplus h_2(t) : (t,x,y) \in \mathcal{Q}_{V_2}\} \\ \widetilde{U}_1 &= \{x \oplus h_1(t) : (t,x,y) \in \mathcal{Q}_{V_2}\}.\end{aligned}$$

Par un raisonnement similaire (c'est à dire que toutes les valeurs dans $\widetilde{V}_1$, respectivement $\widetilde{U}_1$, sont distinctes car τ ne vérifie pas (C-10), respectivement (C-8), $V_1 \cap \widetilde{V}_1 = \emptyset$ car τ ne vérifie pas (C-3), et $U_1 \cap \widetilde{U}_1 = \emptyset$ car τ ne vérifie pas (C-1)), on montre que E_{V_2} est équivalent à α_2 nouvelles équations distinctes sur P_1. Par conséquent,

$$\Pr\left[\mathsf{E}_{V_2} | P_1 \vdash \mathcal{Q}_{P_1}\right] = \frac{1}{(N - q_p)_{\alpha_2}}. \tag{3.4}$$

En combinant (3.3) et (3.4), on obtient

$$\mathsf{p}'(\tau) = \frac{1}{(N - q_p)_{\alpha_1}(N - q_p)_{\alpha_2}}. \tag{3.5}$$

Minoration de $\mathsf{p}''(\tau)$. En conditionnant par les évènements $\mathsf{E}_{U_1} \wedge \mathsf{E}_{V_2} \wedge (P_i \vdash \mathcal{Q}_{P_i}, i = 1, 2)$, P_1 et P_2 sont fixées sur respectivement $q_p + \alpha_2$ et $q_p + \alpha_1$ valeurs. Notre objectif est de minorer le nombre de «valeurs intermédiaires» possibles telles que les évènements $\mathsf{E}_X \wedge \mathsf{E}_Y \wedge \mathsf{E}_0$ soient équivalents à de nouvelles équations distinctes sur P_1 et P_2. Le lecteur peut se référer à la Figure 3.2 lors du comptage.

Commençons par les requêtes de $\mathcal{Q}_X$. Soient $U_1' = \{x \oplus h_1(t) : (t, x, y) \in \mathcal{Q}_X\}$ et $\alpha_1' = |U_1'|$. On constate que

$$\alpha_1' \leq \sum_{\substack{u \in \{0,1\}^n : \\ |X_u| > 1}} 1 \leq \sum_{\substack{u \in \{0,1\}^n : \\ |X_u| > 1}} \frac{|X_u|}{2} = \frac{\beta_1}{2} \leq \frac{\sqrt{q_c}}{2}, \tag{3.6}$$

la dernière inégalité découlant du fait que τ ne vérifie pas la condition (C-13). Par souci de lisibilité, on note, en utilisant un ordre arbitraire,

$$U_1' = \{u_{1,1}', \ldots, u_{1,\alpha_1'}'\}.$$

D'une part, on constate que U_1' est disjoint de U_1 par définition de $\mathcal{Q}_X$, et disjoint de $\widetilde{U}_1$ car sinon τ vérifierait (C-8).

D'autre part, toutes les valeurs $y \oplus h_2(t)$ lorsque (t, x, y) parcourt $\mathcal{Q}_X$ sont distinctes car sinon τ vérifierait (C-4). Soient $V_2' = \{y \oplus h_2(t) : (t, x, y) \in \mathcal{Q}_X\}$ et $\alpha_2' = |V_2'| = |\mathcal{Q}_X|$. On a

$$\alpha_2' = \sum_{i=1}^{\alpha_1'} |X_{u_{1,i}'}| \leq \sum_{\substack{u \in \{0,1\}^n : \\ |X_u| > 1}} |X_u| = \beta_1 \leq \sqrt{q_c}, \tag{3.7}$$

la dernière inégalité étant vraie car τ ne vérifie pas la condition (C-13). Remarquons que V_2' est disjoint de V_2 car sinon τ vérifierait (C-8), et disjoint de $\widetilde{V}_2$ par définition de $\mathcal{Q}_X$.

Soit N_X le nombre d'uplets de valeurs distinctes $(v_{1,i}')_{1 \leq i \leq \alpha_1'}$ dans $\{0,1\}^n \setminus (V_1 \cup \widetilde{V}_1)$ satisfaisant les conditions suivantes :

(i) pour tout i et tout $(t, x, y) \in X_{u_{1,i}'}$, $v_{1,i}' \oplus h_1(t) \oplus h_2(t) \notin (U_2 \cup \widetilde{U}_2)$ (ce qui exclut au plus $(|U_2| + |\widetilde{U}_2|)|X_{u_{1,i}'}| = (q_p + \alpha_1)|X_{u_{1,i}'}|$ valeurs pour $v_{1,i}'$),

(ii) pour tout i et tout $(t, x, y) \in X_{u_{1,i}'}$, $v_{1,i}' \oplus h_1(t) \oplus h_2(t)$ est distinct de toute valeur $v_{1,j}' \oplus h_1(t') \oplus h_2(t')$, pour $j < i$ et $(t', x', y') \in X_{u_{1,j}'}$ (ce qui exclut au plus $|X_{u_{1,i}'}| \cdot \sum_{j=1}^{i-1} |X_{u_{1,j}'}| \leq \alpha_2' |X_{u_{1,i}'}|$ valeurs pour $v_{1,i}'$).

Alors, puisque $|\{0,1\}^n \setminus (V_1 \cup \widetilde{V}_1)| = N - q_p - \alpha_2$, on a

$$N_X \geq \prod_{i=1}^{\alpha_1'} \left(N - q_p - \alpha_2 - (i-1) - (q_p + \alpha_1 + \alpha_2')|X_{u_{1,i}'}| \right). \tag{3.8}$$

On constate que, pour tout uplet de valeurs $(v_{1,i}')$ vérifiant ces conditions, l'ensemble des valeurs $v_{1,i}' \oplus h_1(t) \oplus h_2(t)$ pour $i = 1, \ldots, \alpha_1'$ et $(t, x, y) \in X_{u_{1,i}'}$ est disjoint de $U_2 \cup \widetilde{U}_2$ grâce à la condition (i), et ces valeurs sont deux à deux disjointes grâce à

la condition (ii) et le fait que τ ne vérifie pas (C-5). Par conséquent, lorsque l'on fixe $P_1(u'_{1,i}) = v'_{1,i}$ pour $i = 1, \ldots, \alpha'_1$, l'évènement $\mathsf{TEM}^{P_1,P_2}_{(h_1,h_2)} \vdash \mathcal{Q}_X$ est équivalent à α'_2 «nouvelles» équations distinctes sur P_2.

Pour le reste de la discussion, fixons un tel uplet de valeurs $(v'_{1,i})$, et notons

$$V'_1 = \{v'_{1,1}, \ldots, v'_{1,\alpha'_1}\}$$
$$U'_2 = \{v'_{1,i} \oplus h_1(t) \oplus h_2(t) : i = 1, \ldots, \alpha'_1 \text{ et } (t,x,y) \in X_{u'_{1,i}}\}.$$

Considérons ensuite les requêtes de l'ensemble $\mathcal{Q}_Y$. Soient $V''_2 = \{y \oplus h_2(t) : (t,x,y) \in \mathcal{Q}_V\}$ et $\alpha''_2 = |V''_2|$. On constate que

$$\alpha''_2 \leq \sum_{\substack{v \in \{0,1\}^n: \\ |Y_v|>1}} 1 \leq \sum_{\substack{v \in \{0,1\}^n: \\ |Y_v|>1}} \frac{|Y_v|}{2} = \frac{\beta_2}{2} \leq \frac{\sqrt{q_c}}{2}, \tag{3.9}$$

la dernière inégalité découlant du fait que τ ne vérifie pas (C-14). Par souci de lisibilité, on note, en ordonnant les valeurs de façon arbitraire,

$$V''_2 = \{v''_{2,1}, \ldots, v''_{2,\alpha''_2}\}.$$

Remarquons que, d'une part, V''_2 est disjoint de V_2 par définition de $\mathcal{Q}_Y$, disjoint de $\widetilde{V}_2$, car sinon τ vérifierait la condition (C-7), et disjoint de V'_2 car sinon la condition (C-4) serait satisfaite.

D'autre part, les valeurs $x \oplus h_1(t)$ pour $(t,x,y) \in \mathcal{Q}_Y$ sont distinctes car sinon τ vérifierait la condition (C-4). Soient $U''_1 = \{x \oplus h_1(t) : (t,x,y) \in \mathcal{Q}_Y\}$ et $\alpha''_1 = |U''_1| = |\mathcal{Q}_Y|$. On a

$$\alpha''_1 = \sum_{i=1}^{\alpha''_2} |Y_{v''_{2,i}}| \leq \sum_{\substack{v \in \{0,1\}^n: \\ |Y_v|>1}} |Y_v| = \beta_2 \leq \sqrt{q_c}, \tag{3.10}$$

où la dernière inégalité vient du fait que τ ne vérifie pas la condition (C-14). De plus U''_1 est disjoint de U_1 car sinon τ vérifierait la condition (C-7), disjoint de $\widetilde{U}_1$ par définition de $\mathcal{Q}_Y$, et disjoint de U'_1 car sinon τ vérifierait la condition (C-4).

Soit N_Y le nombre d'uplets de valeurs distinctes $(u''_{2,i})_{1 \leq i \leq \alpha''_2}$ dans $\{0,1\}^n \setminus (U_2 \cup \widetilde{U}_2 \cup U'_2)$ satisfaisant les deux conditions suivantes :

(i) pour tout i et tout $(t,x,y) \in Y_{v''_{2,i}}$, $u''_{2,i} \oplus h_1(t) \oplus h_2(t) \notin (V_1 \cup \widetilde{V}_1 \cup V'_1)$ (ce qui exclut au plus $(|V_1| + |\widetilde{V}_1| + |V'_1|)|Y_{v''_{2,i}}| = (q_p + \alpha_2 + \alpha'_1)|Y_{v''_{2,i}}|$ valeurs pour $u''_{2,i}$),

(ii) pour tout i et tout $(t,x,y) \in Y_{v''_{2,i}}$, $u''_{2,i} \oplus h_1(t) \oplus h_2(t)$ est distinct de toute valeur de la forme $u''_{2,j} \oplus h_1(t') \oplus h_2(t')$ pour $j < i$ et $(t',x',y') \in Y_{v''_{2,j}}$ (ce qui interdit au plus $|Y_{v''_{2,i}}| \cdot \sum_{j=1}^{i-1} |Y_{v''_{2,j}}| \leq \alpha''_1 |Y_{v''_{2,i}}|$ valeurs pour $u''_{2,i}$).

Ainsi, puisque $|\{0,1\}^n \setminus (U_2 \cup \widetilde{U}_2 \cup U'_2)| = N - q_p - \alpha_1 - \alpha'_2$, on a

$$N_Y \geq \prod_{i=1}^{\alpha''_2} \left(N - q_p - \alpha_1 - \alpha'_2 - (i-1) - (q_p + \alpha_2 + \alpha'_1 + \alpha''_1)|Y_{v''_{2,i}}|\right). \tag{3.11}$$

On remarque que, pour tout uplet de valeurs $(u''_{2,i})$ satisfaisant ces deux conditions, l'ensemble des valeurs $u''_{2,i} \oplus h_1(t) \oplus h_2(t)$ pour $i = 1, \ldots, \alpha''_2$ et $(t, x, y) \in Y_{v''_{2,i}}$ est disjoint de $V_1 \cup \widetilde{V}_1 \cup V'_1$ grâce à la condition (i), et ces valeurs sont distinctes grâce à la condition (ii) et le fait que τ ne vérifie pas la condition (C-6). Ainsi, en fixant $P_2^{-1}(v''_{2,i}) = u''_{2,i}$ pour $i = 1, \ldots, \alpha''_2$, l'évènement $\mathsf{TEM}^{P_1,P_2}_{(h_1,h_2)} \vdash \mathcal{Q}_Y$ est équivalent à α''_1 «nouvelles» équations distinctes sur P_1.

À présent, fixons un tel uplet de valeurs $(u''_{2,i})$, et on note

$$\begin{aligned} U''_2 &= \{u''_{2,1}, \ldots, u''_{2,\alpha''_2}\} \\ V''_1 &= \{u''_{2,i} \oplus h_1(t) \oplus h_2(t) : i = 1, \ldots, \alpha''_2 \text{ et } (t, x, y) \in Y_{v''_{2,i}}\}. \end{aligned}$$

Nous devons encore considérer les requêtes de $\mathcal{Q}_0$. Soient

$$\begin{aligned} q'_c &= |\mathcal{Q}_0| = q_c - \alpha_1 - \alpha_2 - \alpha'_2 - \alpha''_1 \\ q'_{p_1} &= |U_1 \cup \widetilde{U}_1 \cup U'_1 \cup U''_1| = q_p + \alpha_2 + \alpha'_1 + \alpha''_1 \\ q'_{p_2} &= |V_2 \cup \widetilde{V}_2 \cup V'_2 \cup V''_2| = q_p + \alpha_1 + \alpha'_2 + \alpha''_2. \end{aligned}$$

On rappelle que m dénote le nombre de tweaks distincts qui apparaissent dans $\mathcal{Q}_C$. Notons $t_1, \ldots, t_m$ ces tweaks (en les ordonnant de façon arbitraire), et pour $i = 1, \ldots, m$, on note $\mathcal{Q}_{0,i}$ le sous-ensemble des requêtes de $\mathcal{Q}_0$ dont le tweak est t_i, et $q'_i = |\mathcal{Q}_{0,i}|$ (certains de ces sous-ensembles peuvent être vides). Remarquons que $\sum_{i=1}^{m} q'_i = q'_c$.

Afin de simplifier le comptage à venir, ordonnons les requêtes de $\mathcal{Q}_0$ de telle façon que les q'_1 premières requêtes soient celles utilisant le tweak t_1, etc. Ainsi, on écrit

$$\mathcal{Q}_0 = \{(t_1, x_{1,1}, y_{1,1}), \ldots, (t_1, x_{1,q'_1}, y_{1,q'_1}), \ldots, (t_m, x_{m,1}, y_{m,1}), \ldots, (t_m, x_{m,q'_m}, y_{m,q'_m})\}.$$

Pour $i = 1, \ldots, m$ et $j = 1, \ldots, q'_i$, soient

$$\begin{aligned} u^{(3)}_{1,i,j} &= x_{i,j} \oplus h_1(t_i) \\ v^{(3)}_{2,i,j} &= y_{i,j} \oplus h_2(t_i). \end{aligned}$$

Constatons que, par définition de $\mathcal{Q}_0$, les $u^{(3)}_{1,i,j}$ sont distincts et n'appartiennent pas à $U_1 \cup \widetilde{U}_1 \cup U'_1 \cup U''_1$, et les $v^{(3)}_{2,i,j}$ sont distincts et n'appartiennent pas à $V_2 \cup \widetilde{V}_2 \cup V'_2 \cup V''_2$.

Soit N_0 le nombre d'uplets de valeurs distinctes $(v^{(3)}_{1,i,j})_{1 \le i \le m, 1 \le j \le q'_i}$ dans $\{0,1\}^n \setminus (V_1 \cup \widetilde{V}_1 \cup V'_1 \cup V''_1)$ satisfaisant les deux conditions suivantes :

(i) pour tout (i, j), $v^{(3)}_{1,i,j} \oplus h_1(t_i) \oplus h_2(t_i) \notin (U_2 \cup \widetilde{U}_2 \cup U'_2 \cup U''_2)$ (ce qui exclut au plus q'_{p_2} valeurs pour $v^{(3)}_{1,i,j}$),

(ii) pour tout $i = 1, \ldots, m$ et tout $j = 1, \ldots, q'_i$, $v^{(3)}_{1,i,j} \oplus h_1(t_i) \oplus h_2(t_i)$ est distinct de toute valeur $v^{(3)}_{1,k,\ell} \oplus h_1(t_k) \oplus h_2(t_k)$ pour $k < i$ et $\ell = 1, \ldots, q'_k$ (ce qui exclut au plus $\sum_{k=1}^{i-1} q'_k$ valeurs pour $v^{(3)}_{1,i,j}$).

Alors, puisque $|\{0,1\}^n \setminus (V_1 \cup \widetilde{V}_1 \cup V_1' \cup V_1'')| = N - q'_{p_1}$, et $v^{(3)}_{1,i,j}$ doivent être différents des $\sum_{k=1}^{i-1} q'_k + (j-1)$ précédentes valeurs $v^{(3)}_{1,k,\ell}$, $k < i$ et $\ell = 1, \ldots, q'_k$, et $v^{(3)}_{1,i,\ell}$, $\ell < j$, on a

$$\begin{aligned} N_0 &\geq \prod_{i=1}^{m} \prod_{j=1}^{q'_i} \left(N - q'_{p_1} - q'_{p_2} - 2\sum_{k=1}^{i-1} q'_k - (j-1) \right) \\ &= \prod_{i=1}^{m} \left(N - q'_{p_1} - q'_{p_2} - 2\sum_{k=1}^{i-1} q'_k \right)_{q'_i} . \end{aligned} \tag{3.12}$$

Pour tout uplet de valeurs $(v^{(3)}_{1,i,j})$ satisfaisant les conditions ci-dessus, l'ensemble des valeurs $v^{(3)}_{1,i,j} \oplus h_1(t_i) \oplus h_2(t_i)$ pour $i = 1, \ldots, m$ et $j = 1, \ldots, q'_i$ est disjoint de $U_2 \cup \widetilde{U}_2 \cup U_2' \cup U_2''$ grâce à la condition (i), et ces valeurs sont distinctes grâce à la condition (ii) et le fait que, pour $i = 1, \ldots, m$, les q'_i valeurs $v^{(3)}_{1,i,j} \oplus h_1(t_i) \oplus h_2(t_i)$ $(1 \leq j \leq q'_i)$ sont nécessairement distincts car les valeurs $v^{(3)}_{1,i,j}$ sont distinctes. Par conséquent, lorsque l'on conditionne sur le fait que P_1 vérifie les q'_c équations $P_1(u^{(3)}_{1,i,j}) = v^{(3)}_{1,i,j}$, l'évènement $\mathsf{TEM}^{P_1,P_2}_{(h_1,h_2)} \vdash \mathcal{Q}_0$ est équivalent à q'_c «nouvelles» équations distinctes sur P_2.

Finalement, nous avons que, pour chacun des (au moins) $N_X \cdot N_Y \cdot N_0$ choix possibles pour les uplets $(v'_{1,i})_{1 \leq i \leq \alpha'_1}$, $(u''_{2,i})_{1 \leq i \leq \alpha''_2}$, et $(\mathbf{v}_{1,i,j})_{1 \leq i \leq m, 1 \leq j \leq q'_i}$ satisfaisant toutes les conditions ci-dessus (ils correspondent aux points rouges dans la Figure 3.2), les évènements $\mathsf{E}_X \wedge \mathsf{E}_Y \wedge \mathsf{E}_0$ sont équivalents à exactement $\alpha'_1 + \alpha''_1 + q'_c$ «nouvelles» équations sur P_1 et exactement $\alpha'_2 + \alpha''_2 + q'_c$ «nouvelles» équations sur P_2 (par nouvelle, on exprime le fait qu'elles ne sont pas imposées par l'évènement $P_1 \vdash \mathcal{Q}_{P_1} \wedge P_2 \vdash \mathcal{Q}_{P_2} \wedge \mathsf{E}_{U_1} \wedge \mathsf{E}_{V_2}$). Par conséquent, on a

$$\mathsf{p}''(\tau) \geq \frac{N_X \cdot N_Y \cdot N_0}{(N - q_p - \alpha_2)_{\alpha'_1 + \alpha''_1 + q'_c} (N - q_p - \alpha_1)_{\alpha'_2 + \alpha''_2 + q'_c}}. \tag{3.13}$$

En rassemblant les équations (3.2), (3.5), and (3.13), on obtient

$$\mathsf{p}(\tau) \geq \frac{N_X \cdot N_Y \cdot N_0}{(N - q_p)_{\alpha_2 + \alpha'_1 + \alpha''_1 + q'_c} (N - q_p)_{\alpha_1 + \alpha'_2 + \alpha''_2 + q'_c}}. \tag{3.14}$$

Finalement, en combinant (2.4) et (3.14), on arrive à

$$\begin{aligned} \frac{\Pr[T_{\mathrm{re}} = \tau]}{\Pr[T_{\mathrm{id}} = \tau]} &\geq \frac{N_X \cdot N_Y \cdot N_0 \cdot \prod_{i=1}^{m} (N)_{q_i}}{(N - q_p)_{\alpha_2 + \alpha'_1 + \alpha''_1 + q'_c} (N - q_p)_{\alpha_1 + \alpha'_2 + \alpha''_2 + q'_c}} \\ &= \underbrace{\frac{N_X}{(N - q_p - \alpha_2)_{\alpha'_1}}}_{R_X} \times \underbrace{\frac{N_Y}{(N - q_p - \alpha_1 - \alpha'_2)_{\alpha''_2}}}_{R_Y} \times \underbrace{\frac{N_0 \cdot \prod_{i=1}^{m} (N)_{q'_i}}{(N - q'_{p_1})_{q'_c} (N - q'_{p_2})_{q'_c}}}_{R_0} \\ &\quad \times \underbrace{\frac{\prod_{i=1}^{m} (N)_{q_i}}{\prod_{i=1}^{m} (N)_{q'_i} (N - q_p)_{\alpha_2} (N - q_p - \alpha_2 - \alpha'_1)_{\alpha''_1} (N - q_p)_{\alpha_1 + \alpha'_2}}}_{R'}. \end{aligned}$$

Il reste à minorer R_X, R_Y, R_0, et R'. En injectant (3.8) dans R_X, on a

$$\begin{aligned} R_X &\geq \frac{\prod_{i=1}^{\alpha'_1}\left(N - q_p - \alpha_2 - (i-1) - (q_p + \alpha_1 + \alpha'_2)|X_{u'_{1,i}}|\right)}{(N - q_p - \alpha_2)_{\alpha'_1}} \\ &= \prod_{i=1}^{\alpha'_1}\left(1 - \frac{(q_p + \alpha_1 + \alpha'_2)|X_{u'_{1,i}}|}{N - q_p - \alpha_2 - (i-1)}\right) \\ &\geq 1 - \frac{(q_p + \alpha_1 + \alpha'_2)\sum_{i=1}^{\alpha'_1}|X_{u'_{1,i}}|}{N - q_p - \alpha_2 - \alpha'_1} \\ &= 1 - \frac{(q_p + \alpha_1 + \alpha'_2)\alpha'_2}{N - q_p - \alpha_2 - \alpha'_1} \\ &\geq 1 - \frac{2\sqrt{q_c}(q_p + 2\sqrt{q_c})}{N}, \end{aligned} \tag{3.15}$$

où, pour obtenir la dernière égalité, on a utilisé le fait que $\alpha_1 \leq \sqrt{q_c}$ car τ est une bonne transcription, $\alpha'_2 \leq \sqrt{q_c}$ par (3.7), et $q_p + \alpha_2 + \alpha'_1 \leq q_p + 2q_c \leq N/2$ par hypothèse.

De même, en injectant (3.11) dans R_Y, on a

$$\begin{aligned} R_Y &\geq \frac{\prod_{i=1}^{\alpha''_2}(N - q_p - \alpha_1 - \alpha'_2 - (i-1) - (q_p + \alpha_2 + \alpha'_1 + \alpha''_1)|Y_{v''_{2,i}}|)}{(N - q_p - \alpha_1 - \alpha'_2)_{\alpha''_2}} \\ &= \prod_{i=1}^{\alpha''_2}\left(1 - \frac{(q_p + \alpha_2 + \alpha'_1 + \alpha''_1)|Y_{v''_{2,i}}|}{N - q_p - \alpha_1 - \alpha'_2 - (i-1)}\right) \\ &\geq 1 - \frac{(q_p + \alpha_2 + \alpha'_1 + \alpha''_1)\sum_{i=1}^{\alpha''_2}|Y_{v''_{2,i}}|}{N - q_p - \alpha_1 - \alpha'_2 - \alpha''_2} \\ &= 1 - \frac{(q_p + \alpha_2 + \alpha'_1 + \alpha''_1)\alpha''_1}{N - q_p - \alpha_1 - \alpha'_2 - \alpha''_2} \\ &\geq 1 - \frac{2\sqrt{q_c}(q_p + 3\sqrt{q_c})}{N}, \end{aligned} \tag{3.16}$$

où, pour la dernière inégalité, on a utilisé le fait que $\alpha_2 \leq \sqrt{q_c}$ puisque τ est une bonne transcription, $\alpha'_1 \leq \sqrt{q_c}$ par (3.6), $\alpha''_1 \leq \sqrt{q_c}$ par (3.10), et $q_p + \alpha_1 + \alpha'_2 + \alpha''_2 \leq q_p + 3q_c \leq N/2$ par hypothèse.

Pour R_0, en utilisant (3.12), on a

$$\begin{aligned} R_0 &\geq \frac{\prod_{i=1}^{m}(N)_{q'_i}\left(N - q'_{p_1} - q'_{p_2} - 2\sum_{k=1}^{i-1}q'_k\right)_{q'_i}}{(N - q'_{p_1})_{q'_c}(N - q'_{p_2})_{q'_c}} \\ &= \prod_{i=1}^{m}\frac{(N)_{q'_i}\left(N - q'_{p_1} - q'_{p_2} - 2\sum_{k=1}^{i-1}q'_k\right)_{q'_i}}{\left(N - q'_{p_1} - \sum_{k=1}^{i-1}q'_k\right)_{q'_i}\left(N - q'_{p_2} - \sum_{k=1}^{i-1}q'_k\right)_{q'_i}} \\ &\geq \prod_{i=1}^{m}\left(1 - \frac{4q'_i\left(q'_{p_1} + \sum_{k=1}^{i-1}q'_k\right)\left(q'_{p_2} + \sum_{k=1}^{i-1}q'_k\right)}{N^2}\right), \end{aligned}$$

où, pour la dernière inégalité, on a utilisé le Lemme 16 avec $a = q'_i$, $b = q'_{p_1} + \sum_{k=1}^{i-1} q'_k$, et $c = q'_{p_2} + \sum_{k=1}^{i-1} q'_k$ (remarquons que $a+b \leq q'_c + q'_{p_1} \leq q_c + q_p + \alpha'_1 \leq q_p + 2q_c \leq N/2$ par hypothèse, et de même $a + c \leq N/2$).

Constatons que, par définition de q'_c, q'_{p_1} et q'_{p_2}, on a

$$q'_{p_1} + \sum_{k=1}^{i-1} q'_k \leq q'_{p_1} + q'_c \leq q_p + q_c + \alpha'_1 \leq q_p + 2q_c,$$
$$q'_{p_2} + \sum_{k=1}^{i-1} q'_k \leq q'_{p_2} + q'_c \leq q_p + q_c + \alpha''_2 \leq q_p + 2q_c,$$

où nous avons utilisé les inégalités (3.6) and (3.9), de sorte que nous arrivons à l'inégalité suivante

$$R_0 \geq 1 - \frac{4q_c(q_p + 2q_c)^2}{N^2}. \tag{3.17}$$

Finalement, on a

$$\begin{aligned} R' &= \frac{\prod_{i=1}^{m}(N - q'_i)_{q_i - q'_i}}{(N - q_p)_{\alpha_2}(N - q_p - \alpha_2 - \alpha'_1)_{\alpha''_1}(N - q_p)_{\alpha_1 + \alpha'_2}} \\ &\geq \frac{\prod_{i=1}^{m}(N - q'_i)_{q_i - q'_i}}{N^{\alpha_1 + \alpha_2 + \alpha'_2 + \alpha''_1}} \\ &\geq \frac{(N - q_c)^{\sum_{i=1}^{m} q_i - q'_i}}{N^{\alpha_1 + \alpha_2 + \alpha'_2 + \alpha''_1}} \\ &= \frac{(N - q_c)^{q_c - q'_c}}{N^{\alpha_1 + \alpha_2 + \alpha'_2 + \alpha''_1}} \\ &\geq 1 - \frac{4q_c^{3/2}}{N}, \end{aligned} \tag{3.18}$$

car

$$q_c - q'_c = \alpha_1 + \alpha_2 + \alpha'_2 + \alpha''_1 \leq 4\sqrt{q_c}.$$

On conclut la preuve en rassemblant (3.15), (3.16), (3.17) et (3.18). □

Conclusion

Nous pouvons maintenant prouver le théorème 5. En combinant les lemmes 1, 15 et 17, on a

$$\begin{aligned} \mathbf{Adv}^{\mathrm{cca}}_{\mathsf{TEM}[n,2,\mathbf{f}_{\mathcal{H}}]}(q_c, q_p) &\leq \frac{3q_c q_p^2}{N^2} + 2\varepsilon^2 q_c^3 + \frac{\varepsilon q_c^2 q_p}{N} + \frac{2\sqrt{q_c} q_p}{N} + 4\varepsilon q_c^{3/2} \\ &\quad + \frac{4q_c(q_p + 2q_c)^2}{N^2} + \frac{14q_c^{3/2} + 4\sqrt{q_c} q_p}{N} \\ &= \frac{7q_c q_p^2}{N^2} + \frac{16q_c^2 q_p}{N^2} + \frac{6\sqrt{q_c} q_p}{N} + \frac{\varepsilon q_c^2 q_p}{N} + 2\varepsilon^2 q_c^3 + 4\varepsilon q_c^{3/2} \\ &\quad + \frac{16q_c^3}{N^2} + \frac{14q_c^{3/2}}{N} \end{aligned}$$

$$\leq \frac{7q_c q_p^2}{N^2} + \frac{16q_c^2 q_p}{N^2} + \frac{6\sqrt{q_c}q_p}{N} + \frac{\varepsilon q_c^2 q_p}{N} + 6\varepsilon q_c^{3/2} + \frac{30q_c^{3/2}}{N},$$

où, pour la dernière inégalité, on a utilisé l'hypothèse que $q_c \leq \min\{N^{2/3}, \varepsilon^{-2/3}\}$. Comme le résultat est trivialement vrai lorsque $q_c q_p^2 > N^2$, on peut supposer que $q_c q_p^2 \leq N^2$, ainsi $q_c q_p^2/N^2 \leq \sqrt{q_c}q_p/N$. De plus, comme $q_c \leq N^{2/3}$, on a $q_c^2/N^2 \leq \sqrt{q_c}/N$ et $q_c^2/N \leq \sqrt{q_c}$, ce qui achève la preuve du théorème 5.

3.4 Preuve de sécurité asymptotique pour r tours

3.4.1 Objectif

Dans cette section, nous allons étudier la sécurité de la construction $\mathsf{TEM}[r, n, \mathbf{f}_{\mathcal{H}}]$ pour un nombre arbitraire r de tours. Jusqu'à présent, grâce à la technique des coefficients H, nous avons pu prouver que cette construction est sûre tant que le nombre de requêtes de l'adversaire reste petit devant $2^{rn/(r+1)}$ pour $r = 1, 2$. Nous conjecturons que, comme dans le cas du schéma d'Even-Mansour classique, ce résultat reste vrai pour $r \geq 3$, toutefois donner une preuve de ce résultat reste un problème ouvert difficile. La preuve de sécurité dans le cas non-paramétrable [CS14] ne peut notamment pas se transposer aisément au cas paramétrable : en effet, dans la preuve de Chen et Steinberger, les valeurs qui sont XORée entre les tours ne dépendent pas des requêtes. Ainsi, il ne peut se produire aucune collision entre des valeurs intermédiaires rencontrées lors du calcul de requêtes de chiffrement et les seules collisions à considérer sont celles qui lient le chiffrement ou le déchiffrement d'une requête à la construction et une chaîne de requêtes aux oracles de permutations. Cependant, dans le cas de notre construction, comme nous avons pu le constater, ces deux types de collisions se produisent avec une probabilité proche de 1, ce qui rend l'analyse des bonnes transcriptions d'autant plus complexe.

Nous allons plutôt donner une analyse asymptotique de cette construction, quand le nombre de tour augmente, grâce à la technique probabiliste du couplage [MRS09, HR10]. Après avoir présenté brièvement son fonctionnement, nous l'utiliserons afin d'étudier la sécurité de notre construction face aux adversaires ncpa, la technique du couplage n'étant pas adaptée à l'étude d'adversaires adaptatifs. Dans ce contexte, un adversaire ncpa fonctionne en deux phases : au cours de la première phase, il n'interagit qu'avec les permutations internes, de façon adaptative et dans les deux directions ; dans la seconde phase, il fournit un uplet de requêtes non-adaptatives à clairs choisis à l'oracle de construction, puis reçoit les réponses correspondantes (ces requêtes peuvent dépendre des réponses reçues lors de la première phase, mais l'intégralité des requêtes de la seconde phase doit être choisie avant toute réponse de l'oracle de construction). Cette partie combine les arguments de [LPS12], qui concernaient le schéma d'Even-Mansour classique et ceux de [LS14], dans le cadre de la construction LRW^2 itérée. Nous utiliserons ensuite le lemme 19, qui peut être vu comme une réciproque partielle au lemme des coefficients H, pour démontrer la sécurité de la construction TEM face aux adversaires adaptatifs à clairs et chiffrés choisis jusqu'à environ $2^{rn/(r+2)}$ requêtes.

3.4.2 Technique employée

Soient μ et ν deux lois de probabilité sur un même espace mesurable $(E, \mathcal{E})$. Un *couplage* de μ et ν est une loi de probabilité λ sur $E \times E$ tel que pour tout $x \in E$, $\sum_{y \in E} \lambda(x, y) = \mu(x)$ et pour tout $y \in E$, $\sum_{x \in E} \lambda(x, y) = \nu(y)$. En d'autres termes, λ est une loi dont les marginales sont respectivement μ et ν. Le résultat fondamental de la technique du couplage est le suivant.

Lemme 18 (Lemme du couplage)**.** *Soient μ et ν deux lois de probabilité sur un espace d'état fini Ω, et soient λ un couplage de μ et ν, et $(X, Y) \sim \lambda$ (c'est-à-dire que le couple (X, Y) est une variable aléatoire distribuée selon la loi λ). Alors $\|\mu - \nu\| \leq \Pr[X \neq Y]$.*

Démonstration. Soient λ un couplage de μ et ν et $(X, Y) \sim \lambda$, c'est-à-dire que, pour tout $x, y \in \Omega$, on a

$$\begin{aligned}
\lambda(x, y) &= \Pr\left[X = x, Y = y\right], \\
\mu(x) &= \Pr\left[X = x\right] = \sum_{z \in \Omega} \lambda(x, z) \\
\nu(y) &= \Pr\left[Y = y\right] = \sum_{z \in \Omega} \lambda(z, y).
\end{aligned}$$

En particulier, pour tout $z \in \Omega$, on a

$$\lambda(z, z) = \Pr\left[X = z, Y = z\right] \leq \min\{\Pr\left[X = z\right], \Pr\left[Y = z\right]\} \leq \min\{\mu(z), \nu(z)\}.$$

D'où

$$\begin{aligned}
\Pr\left[X \neq Y\right] &= 1 - \Pr\left[X = Y\right] \\
&= 1 - \sum_{z \in \Omega} \Pr\left[X = z, Y = z\right] \\
&= \sum_{z \in \Omega} \mu(z) - \sum_{z \in \Omega} \lambda(z, z) \\
&\geq \sum_{z \in \Omega} \left(\mu(z) - \min\{\mu(z), \nu(z)\}\right) \\
&= \sum_{\substack{z \in \Omega \\ \mu(z) > \nu(z)}} \left(\mu(z) - \nu(z)\right) \\
&= \|\mu - \nu\|.
\end{aligned}$$

□

Ce lemme permet de majorer simplement une distance statistique par la probabilité que deux variables aléatoires soient différentes. Il est important de noter que l'indépendance de X et Y n'est pas requise : ceci permet de pouvoir corréler astucieusement les deux variables aléatoires afin de rendre le calcul de la probabilité qu'elles soient différentes le plus simple possible.

Signalons également que cette technique est en fait optimale : quelles que soient les lois de probabilité μ et ν, il existe un couplage λ tel que l'inégalité du lemme soit une égalité (voir [Lin02] pour une preuve de ce résultat).

3.4.3 Préliminaires et Notations

Fixons un entier naturel $q \leq N$. Étant donné un uplet $\mathbf{t} = (t_1, \ldots, t_q) \in \mathcal{T}^q$, nous noterons $\Omega_{\mathbf{t}} \subset (\{0,1\}^n)^q$ l'ensemble des entrées possibles $\mathbf{x} = (x_1, \ldots, x_q) \in (\{0,1\}^n)^q$ telles que toutes les paires (t_i, x_i) sont deux à deux distinctes, c'est-à-dire

$$\Omega_{\mathbf{t}} = \{\mathbf{x} := (x_1, \ldots, x_q) \in (\{0,1\}^n)^q : \forall i \neq j, (x_i, t_i) \neq (x_j, t_j)\}.$$

Pour l'analyse des attaques cca, nous utiliserons le lemme suivant.

Lemme 19. *Soient Ω un ensemble fini d'évènement et μ^* la distribution de probabilité uniforme sur Ω. Soit μ une distribution de probabilité sur Ω telle que $\|\mu - \mu^*\| \leq \varepsilon$. Alors il existe un ensemble $S \subset \Omega$ tel que :*
— $|S| \geq (1 - \sqrt{\varepsilon})|\Omega|$,
— $\forall x \in S$, $\mu(x) \geq (1 - \sqrt{\varepsilon})\mu^*(x)$.

Démonstration. La preuve de ce résultat est classique, voir par exemple [LPS12]. On définit S par

$$S = \{x \in \Omega \,:\, \mu(x) \geq (1 - \sqrt{\varepsilon})\nu(x)\}$$

et montrons par l'absurde que $|S| \geq (1 - \sqrt{\varepsilon})|\Omega|$. Supposons alors que $|S| < (1-\sqrt{\varepsilon})|\Omega|$, ou de manière équivalente que $|\overline{S}| > \sqrt{\varepsilon}|\Omega|$ où $\overline{S}$ désigne le complémentaire de S dans Ω. Ceci entraîne que $\nu(\overline{S}) > \sqrt{\varepsilon}$. Par définition, pour tout $x \in \overline{S}$, $\nu(x) - \mu(x) > \sqrt{\varepsilon}\nu(x)$. Ainsi

$$\nu(\overline{S}) - \mu(\overline{S}) > \sqrt{\varepsilon}\nu(\overline{S}) > (\sqrt{\varepsilon})^2 = \varepsilon,$$

ce qui entre en contradiction avec $||\mu - \nu|| \leq \varepsilon$. □

3.4.4 Analyse de sécurité face aux adversaires non adaptatifs

Tout d'abord, étudions les adversaires non-adaptatifs à clair choisi (ncpa). En utilisant la technique du couplage, nous prouverons le théorème suivant.

Théorème 6. *Soient n, r, q_c, q_p des entiers naturels. Alors on a :*

$$\mathbf{Adv}^{\mathrm{ncpa}}_{\mathsf{TEM}[n,r,\mathbf{f}_{\mathcal{H}}]}(q_c, q_p) \leq 2^r \frac{q_c(N\varepsilon q_c + q_p)^r}{N^r}.$$

En utilisant une famille de fonctions ε-AXU avec $\varepsilon \simeq 2^{-n}$, on constate que la construction assure la sécurité jusqu'à approximativement $2^{\frac{rn}{r+1}}$ requêtes faces aux adversaires ncpa.

Le point crucial de cette preuve sera la majoration de la distance statistique entre la distribution des sorties du schéma d'Even-Mansour paramétrable *conditionnées par des informations partielles sur les permutations internes* (concrètement $P_i \vdash \mathcal{Q}_{P_i}$ pour $i = 1, \ldots, r$) et la distribution uniforme sur $\Omega_{\mathbf{t}}$. Nous introduisons les notations et les définitions suivantes.

Soient q_c, q_p des entiers naturels, et fixons un (q_c, q_p)-distingueur ncpa noté $\mathcal{D}$. Un tel distingueur commence par ses q_p requêtes aux permutations internes $(P_1, \ldots, P_r)$. Soit $\mathcal{Q}_{\mathbf{P}} = (\mathcal{Q}_{P_1}, \ldots, \mathcal{Q}_{P_r})$ la transcription résultant de cette interaction. On dit qu'une transcription $\mathcal{Q}_{\mathbf{P}}$ est *atteignable* s'il existe un uplet de permutations $\mathbf{P}$ tel que l'interaction de $\mathcal{D}$ avec $\mathbf{P}$ donne $\mathcal{Q}_{\mathbf{P}}$. On rappelle qu'on note $\mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}}$ l'évènement $\wedge_{i=1}^{r}(P_i \vdash \mathcal{Q}_{P_i})$.

Définition 9. *Fixons une transcription atteignable des requêtes $\mathcal{Q}_{\mathbf{P}} = (\mathcal{Q}_{P_1}, \ldots, \mathcal{Q}_{P_r})$ résultant de l'interaction adaptative du distingueur avec les permutations internes au cours de la première phase de l'attaque. Pour $\mathbf{t} = (t_1, \ldots, t_{q_c}) \in \mathcal{T}^{q_c}$ et $\mathbf{x} = (x_1, \ldots, x_{q_c}) \in \Omega_{\mathbf{t}}$, on note $\mu_{\mathbf{t},\mathbf{x},\mathcal{Q}_{\mathbf{P}}}$ la distribution de l'uplet*

$$\mathsf{TEM}_{\mathbf{k}}^{\mathbf{P}}(\mathbf{t}, \mathbf{x}) \stackrel{\text{def}}{=} \left(\mathsf{TEM}_{\mathbf{k}}^{\mathbf{P}}(t_1, x_1), \ldots, \mathsf{TEM}_{\mathbf{k}}^{\mathbf{P}}(t_{q_c}, x_{q_c})\right)$$

conditionnée par l'évènement $\mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}}$ (c'est-à-dire quand la clé $\mathbf{k} = (k_1, \ldots, k_r)$ est uniformément aléatoire et les permutations $\mathbf{P} = (P_1, \ldots, P_r)$ sont uniformément aléatoires parmi les permutations satisfaisant $\wedge_{i=1}^{r}(P_i \vdash \mathcal{Q}_{P_i})$). On note également $\mu_{\mathbf{t}}^*$ la distribution uniforme sur $\Omega_{\mathbf{t}}$. ◇

Le lemme suivant majore l'avantage de tout distingueur ncpa par le maximum sur tous les uplets de valeurs $\mathbf{t} \in \mathcal{T}^{q_c}$ et $\mathbf{x} \in \Omega_{\mathbf{t}}$ de la distance statistique entre $\mu_{\mathbf{t},\mathbf{x},\mathcal{Q}_{\mathbf{P}}}$ et $\mu_{\mathbf{t}}^*$. Il s'agit d'une extension directe de [LPS12, Lemme 4].

Lemme 20. *Soient q_c, q_p des entiers naturels. On suppose qu'il existe α tel que, pour toute transcription atteignable des requêtes de la première phase de l'attaque $\mathcal{Q}_{\mathbf{P}}$ et tout $\mathbf{t} \in \mathcal{T}^{q_c}, \mathbf{x} \in \Omega_{\mathbf{t}}$, on a*

$$\|\mu_{\mathbf{t},\mathbf{x},\mathcal{Q}_{\mathbf{P}}} - \mu_{\mathbf{t}}^*\| \leq \alpha.$$

Alors $\mathbf{Adv}^{\text{ncpa}}_{\mathsf{TEM}[n,r,\mathbf{f}_{\mathcal{H}}]}(q_c, q_p) \leq \alpha$.

Démonstration. Fixons un (q_c, q_p)-distingueur ncpa noté $\mathcal{D}$. Soit $\mathcal{Q}_{\mathbf{P}}$ la transcription de l'interaction de $\mathcal{D}$ avec les permutations internes au cours de la première phase de l'attaque. Notons Γ l'ensemble des transcriptions atteignables. Le nombre de ces transcriptions est exactement

$$|\Gamma| = ((N)_{q_p})^r. \tag{3.19}$$

Ceci peut être facilement vu de la façon suivante. La première requête de $\mathcal{D}$ est identique pour toutes les exécutions. Supposons *spg* que c'est une requête à P_1. Il y a exactement N réponses possibles. La requête suivante est déterminée par la réponse reçue à la première requête. S'il s'agit à nouveau d'une requête à P_1, il y a maintenant $N-1$ réponses possibles, alors que, s'il s'agit d'une requête à P_i, $i \neq 1$, il y a N réponses possibles. Cet argument peut être facilement étendu par récurrence pour obtenir l'affirmation ci-dessus.

Les uplets de requêtes non-adaptatives $((t_1, x_1), \ldots, (t_{q_c}, x_{q_c}))$ de $\mathcal{D}$ à l'oracle de construction dépendent de façon déterministe de la transcription $\mathcal{Q}_{\mathbf{P}}$ de la première phase de l'attaque. Notons $\mathbf{t}(\mathcal{Q}_{\mathbf{P}}) = (t_1, \ldots, t_{q_c})$ et $\mathbf{x}(\mathcal{Q}_{\mathbf{P}}) = (x_1, \ldots, x_{q_c})$. La sortie de $\mathcal{D}$ dépend alors de façon déterministe de $\mathcal{Q}_{\mathbf{P}}$ et des réponses $\mathbf{y} = (y_1, \ldots, y_{q_c})$ de l'oracle de construction à l'uplet de requêtes $((t_1, x_1), \ldots, (t_{q_c}, x_{q_c}))$. Pour toute

transcription atteignable $\mathcal{Q}_{\mathbf{P}}$, on note $\Sigma_{\mathcal{Q}_{\mathbf{P}}}$ l'ensemble des uplets $\mathbf{y}$ tels que $\mathcal{D}$ renvoie 1 lorsqu'il reçoit les réponses $\mathbf{y}$ à ses requêtes à l'oracle de construction. Notons

$$\mathsf{p}_{\mathrm{id}} = \Pr\left[\widetilde{P} \leftarrow_{\$} \mathsf{TP}(\mathcal{T}, n), \mathbf{P} \leftarrow_{\$} (\mathsf{P}(n))^r : \mathcal{D}^{\widetilde{P},\mathbf{P}} = 1\right]$$
$$\mathsf{p}_{\mathrm{re}} = \Pr\left[\mathbf{k} \leftarrow_{\$} \mathcal{K}^r, \mathbf{P} \leftarrow_{\$} (\mathsf{P}(n))^r : \mathcal{D}^{\mathsf{TEM}^{\mathbf{P}}_{\mathbf{k}},\mathbf{P}} = 1\right].$$

Alors, par définition, on a

$$\begin{aligned}
\mathsf{p}_{\mathrm{id}} &= \sum_{\mathcal{Q}_{\mathbf{P}} \in \Gamma} \sum_{\mathbf{y} \in \Sigma_{\mathcal{Q}_{\mathbf{P}}}} \Pr\Big[\widetilde{P} \leftarrow_{\$} \mathsf{TP}(\mathcal{T}, n), \mathbf{P} \leftarrow_{\$} (\mathsf{P}(n))^r : \mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}} \\
&\qquad\qquad \wedge \widetilde{P}(\mathbf{t}(\mathcal{Q}_{\mathbf{P}}), \mathbf{x}(\mathcal{Q}_{\mathbf{P}})) = \mathbf{y}\Big] \\
&= \sum_{\mathcal{Q}_{\mathbf{P}} \in \Gamma} \sum_{\mathbf{y} \in \Sigma_{\mathcal{Q}_{\mathbf{P}}}} \Pr\left[\widetilde{P} \leftarrow_{\$} \mathsf{TP}(\mathcal{T}, n) : \widetilde{P}(\mathbf{t}(\mathcal{Q}_{\mathbf{P}}), \mathbf{x}(\mathcal{Q}_{\mathbf{P}})) = \mathbf{y}\right] \qquad (3.20) \\
&\quad \times \Pr\left[\mathbf{P} \leftarrow_{\$} (\mathsf{P}(n))^r : \mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}}\right] \\
&= \sum_{\mathcal{Q}_{\mathbf{P}} \in \Gamma} \Pr\left[\mathbf{P} \leftarrow_{\$} (\mathsf{P}(n))^r : \mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}}\right] \cdot \mu^*_{\mathbf{t}(\mathcal{Q}_{\mathbf{P}})}(\Sigma_{\mathcal{Q}_{\mathbf{P}}}). \qquad (3.21)
\end{aligned}$$

De plus, on a :

$$\begin{aligned}
\mathsf{p}_{\mathrm{re}} &= \sum_{\mathcal{Q}_{\mathbf{P}} \in \Gamma} \sum_{\mathbf{y} \in \Sigma_{\mathcal{Q}_{\mathbf{P}}}} \Pr\left[(\mathbf{k}, \mathbf{P}) \leftarrow_{\$} \mathcal{K}^r \times \mathsf{P}(n)^r : \mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}} \wedge \mathsf{TEM}^{\mathbf{P}}_{\mathbf{k}}(\mathbf{t}(\mathcal{Q}_{\mathbf{P}}), \mathbf{x}(\mathcal{Q}_{\mathbf{P}})) = \mathbf{y}\right] \\
&= \sum_{\mathcal{Q}_{\mathbf{P}} \in \Gamma} \sum_{\mathbf{y} \in \Sigma_{\mathcal{Q}_{\mathbf{P}}}} \Pr\left[\mathbf{P} \leftarrow_{\$} (\mathsf{P}(n))^r : \mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}}\right] \cdot \mu_{\mathbf{t}(\mathcal{Q}_{\mathbf{P}}), \mathbf{x}(\mathcal{Q}_{\mathbf{P}}), \mathcal{Q}_{\mathbf{P}}}(\mathbf{y}) \\
&= \sum_{\mathcal{Q}_{\mathbf{P}} \in \Gamma} \Pr\left[\mathbf{P} \leftarrow_{\$} (\mathsf{P}(n))^r : \mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}}\right] \cdot \mu_{\mathbf{t}(\mathcal{Q}_{\mathbf{P}}), \mathbf{x}(\mathcal{Q}_{\mathbf{P}}), \mathcal{Q}_{\mathbf{P}}}(\Sigma_{\mathcal{Q}_{\mathbf{P}}}). \qquad (3.22)
\end{aligned}$$

Par définition et en utilisant (3.21) et (3.22), on a

$$\begin{aligned}
\mathbf{Adv}(\mathcal{D}) &= |\mathsf{p}_{\mathrm{id}} - \mathsf{p}_{\mathrm{re}}| \\
&\leq \sum_{\mathcal{Q}_{\mathbf{P}} \in \Gamma} \Pr\left[\mathbf{P} \leftarrow_{\$} (\mathsf{P}(n))^r : \mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}}\right] \cdot \left|\mu^*_{\mathbf{t}(\mathcal{Q}_{\mathbf{P}})}(\Sigma_{\mathcal{Q}_{\mathbf{P}}}) - \mu_{\mathbf{t}(\mathcal{Q}_{\mathbf{P}}), \mathbf{x}(\mathcal{Q}_{\mathbf{P}}), \mathcal{Q}_{\mathbf{P}}}(\Sigma_{\mathcal{Q}_{\mathbf{P}}})\right| \\
&\leq \sum_{\mathcal{Q}_{\mathbf{P}} \in \Gamma} \Pr\left[\mathbf{P} \leftarrow_{\$} (\mathsf{P}(n))^r : \mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}}\right] \cdot \|\mu^*_{\mathbf{t}(\mathcal{Q}_{\mathbf{P}})} - \mu_{\mathbf{t}(\mathcal{Q}_{\mathbf{P}}), \mathbf{x}(\mathcal{Q}_{\mathbf{P}}), \mathcal{Q}_{\mathbf{P}}}\| \\
&\leq \alpha \sum_{\mathcal{Q}_{\mathbf{P}} \in \Gamma} \Pr\left[\mathbf{P} \leftarrow_{\$} (\mathsf{P}(n))^r : \mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}}\right] \\
&\leq \alpha,
\end{aligned}$$

où, pour la dernière inégalité, on a utilisé

$$\Pr[\mathbf{P} \leftarrow_{\$} (\mathsf{P}(n))^r : \mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}}] = \frac{1}{((N)_{q_p})^r} = \frac{1}{|\Gamma|}.$$

□

Nous allons maintenant déterminer un majorant adapté α pour $\|\mu_{\mathbf{t},\mathbf{x},\mathcal{Q}_{\mathbf{P}}} - \mu^*_{\mathbf{t}}\|$, ce qui nous permettra d'appliquer le lemme 20.

Lemme 21. *Soient q_c, q_p des entiers naturels. Fixons une transcription atteignable quelconque des requêtes de la première phase de l'attaque $\mathcal{Q}_{\mathbf{P}}$ et soient $\mathbf{t} \in \mathcal{T}^{q_c}, \mathbf{x} \in \Omega_{\mathbf{t}}$. Alors :*

$$\|\mu_{\mathbf{t},\mathbf{x},\mathcal{Q}_{\mathbf{P}}} - \mu^*_{\mathbf{t}}\| \leq q_c \left(2q_c\varepsilon + \frac{2q_p}{N}\right)^r.$$

Démonstration. Fixons une transcription atteignable quelconque $\mathcal{Q}_{\mathbf{P}} = (\mathcal{Q}_{P_1}, \ldots, \mathcal{Q}_{P_r})$ et $\mathbf{t} \in \mathcal{T}^{q_c}, \mathbf{x} \in \Omega_{\mathbf{t}}$, où $\mathbf{t} = (t_1, \ldots, t_{q_c})$ et $\mathbf{x} = (x_1, \ldots, x_{q_c})$. Pour chaque $l \in \{0, \ldots, q_c\}$, soit $\mathbf{z} = (z_1, \ldots, z_{q_c})$ un uplet de requêtes tel que $z_i = x_i$ pour $i \leq l$, et z_i est uniformément aléatoire dans $\{0,1\}^n \setminus \{z_j | t_j = t_i, j < i\}$ pour $i > l$. Notons ν_l la distribution de $\mathsf{TEM}^{\mathbf{P}}_{\mathbf{k}}(\mathbf{t}, \mathbf{z})$, conditionnée par l'évènement $\mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}}$. Notons que $\nu_0 = \mu^*_{\mathbf{t}}$ puisque, pour $l = 0$, l'uplet d'entrées est uniformément aléatoire dans $\Omega_{\mathbf{t}}$, et $\nu_{q_c} = \mu_{\mathbf{t},\mathbf{x},\mathcal{Q}_{\mathbf{P}}}$. Par conséquent, on a :

$$\|\mu_{\mathbf{t},\mathbf{x},\mathcal{Q}_{\mathbf{P}}} - \mu^*_{\mathbf{t}}\| = \|\nu_{q_c} - \nu_0\| \leq \sum_{l=0}^{q_c-1} \|\nu_{l+1} - \nu_l\|. \tag{3.23}$$

Il reste à majorer la distance statistique entre ν_{l+1} et ν_l, pour tout $l \in \{0, \ldots, q_c - 1\}$. Dans ce but, nous allons construire un couplage convenable des deux distributions. On note qu'il suffit de considérer les $l+1$ premiers éléments des deux uplets de sorties puisque, pour les deux distributions, la i-ème entrée pour $i > l+1$ est aléatoire. En d'autres termes, $\|\nu_{l+1} - \nu_l\| = \|\nu'_{l+1} - \nu'_l\|$, où ν'_{l+1} et ν'_l sont les distributions respectives des $l+1$ premières sorties du schéma. Pour définir le couplage de ν'_{l+1} et ν'_l, on considère le schéma d'Even-Mansour paramétrable $\mathsf{TEM}^{\mathbf{P}}_{\mathbf{k}}$, où $\mathbf{P}$ satisfait $\mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}}$, qui reçoit comme entrées $\mathbf{x}' = (x_1, \ldots, x_{l+1})$ et $\mathbf{t}' = (t_1, \ldots, t_{l+1})$, de sorte que $\mathsf{TEM}^{\mathbf{P}}_{\mathbf{k}}(\mathbf{t}', \mathbf{x}')$ est distribué selon ν'_{l+1}. Nous allons construire un second schéma d'Even-Mansour paramétrable $\mathsf{TEM}^{\mathbf{P}'}_{\mathbf{k}'}$, recevant $\mathbf{z}' = (z_1, \ldots, z_{l+1})$ et $\mathbf{t}' = (t_1, \ldots, t_{l+1})$, satisfaisant les propriétés suivantes :

(i) $z_i = x_i$ pour $i = 1, \ldots, l$, et z_{l+1} est uniformément aléatoire dans $\{0,1\}^n \setminus \{x_j | t_j = t_i, j < i\}$;

(ii) pour $i = 1, \ldots, l+1$, si les sorties de la j-ème permutation interne dans le calcul de $\mathsf{TEM}^{\mathbf{P}}_{\mathbf{k}}(t_i, x_i)$ et $\mathsf{TEM}^{\mathbf{P}}_{\mathbf{k}'}(t_i, z_i)$ sont égales, alors cette égalité reste vraie pour toutes les permutations internes suivantes ;

(iii) $\mathbf{P}'$ est uniformément aléatoire parmi les uplets de permutations satisfaisant $\mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}}$ et $\mathbf{k}'$ est uniformément aléatoire dans $\mathcal{K}^r$.

Notons que les mêmes *tweaks* sont utilisés pour les deux schémas. Les propriétés (i) et (iii) assurent que $\mathsf{TEM}^{\mathbf{P}'}_{\mathbf{k}'}(\mathbf{t}', \mathbf{z}')$ est distribué selon ν'_l. Insistons sur le fait que $(\mathbf{P}', \mathbf{k}')$ *ne sera pas* indépendant de $(\mathbf{P}, \mathbf{k})$, toutefois ce n'est pas nécessaire pour pouvoir appliquer le lemme du couplage. Il suffit simplement de vérifier que $(\mathbf{P}, \mathbf{k})$ et $(\mathbf{P}', \mathbf{k}')$ aient les distributions marginales souhaitées.

Notation. Pour $i = 1, \ldots, r$, on note

$$U_i = \{u_i | (u_i, v_i) \in \mathcal{Q}_{P_i}\},$$
$$V_i = \{v_i | (u_i, v_i) \in \mathcal{Q}_{P_i}\}.$$

Pour $i = 1, \ldots, l+1$ et $j = 1, \ldots, r$, on définit également x_i^j comme la sortie du j-ème tour lors du calcul de $\mathsf{TEM}_{\mathbf{k}}^{\mathbf{P}}(t_i, x_i)$, et de même z_i^j désigne la sortie du j-ème tour lors du calcul de $\mathsf{TEM}_{\mathbf{k}'}^{\mathbf{P}'}(t_i, z_i)$, i.e.,

$$\begin{cases} x_i^0 &= x_i, \\ z_i^0 &= z_i, \\ x_i^j &= H_{k_j}(t_i) \oplus P_j(H_{k_j}(t_i) \oplus x_i^{j-1}), \\ z_i^j &= H_{k'_j}(t_i) \oplus P'_j(H_{k'_j}(t_i) \oplus z_i^{j-1}). \end{cases} \tag{3.24}$$

Décrivons maintenant comment construire le second schéma d'Even-Mansour paramétrable. Tout d'abord, il utilise les mêmes clés que le schéma original, c'est-à-dire $\mathbf{k}' = \mathbf{k} = (k_1, \ldots, k_r)$. Afin de construire les permutations $\mathbf{P}'$ (sur les points rencontrés lors du calcul de $\mathsf{TEM}_{\mathbf{k}}^{\mathbf{P}'}(\mathbf{t}', \mathbf{z}')$), on compare les calculs de $\mathsf{TEM}_{\mathbf{k}}^{\mathbf{P}}(t_i, x_i)$ et $\mathsf{TEM}_{\mathbf{k}}^{\mathbf{P}}(t_i, z_i)$ pour $i = 1, \ldots, l+1$.

Le Couplage des l premières requêtes. Pour chaque $i \leq l$, les i-èmes requêtes x_i^0 et z_i^0 sont égales par définition. Considérant le système (3.24), on définit $P'_{j+1}(x_i^j \oplus H_{k_{j+1}}(t_i)) = P_{j+1}(x_i^j \oplus H_{k_{j+1}}(t_i))$ pour chaque $i \leq l$ et $i < r$. Ceci entraîne que les l premières sorties $(x_1^r, \ldots, x_l^r)$ et $(z_1^r, \ldots, z_l^r)$ sont égales.

Le Couplage de la $(l+1)$-ème requête. Pour chaque $j = 0, \ldots, r-1$, on définit le couplage pour la $l+1$-ème requête comme suit :

(1) si $z_{l+1}^j \oplus H_{k_{j+1}}(t_{l+1}) \in U_{j+1}$ ou s'il existe $i \leq l$ tel que $z_{l+1}^j \oplus H_{k_{j+1}}(t_{l+1}) = z_i^j \oplus H_{k_{j+1}}(t_i)$, alors $P'_{j+1}(H_{k_{j+1}}(t_{l+1}) \oplus z_{l+1}^j)$ est déjà déterminé ; à moins que le couplage de z_{l+1}^j et x_{l+1}^j ait eu lieu à un tour précédent, il est impossible de le faire à ce tour ;

(2) si $z_{l+1}^j \oplus H_{k_{j+1}}(t_{l+1}) \notin U_{j+1}$ et $z_{l+1}^j \oplus H_{k_{j+1}}(t_{l+1}) \neq z_i^j \oplus H_{k_{j+1}}(t_i)$ pour $i \leq l$, alors :

   (a) si $x_{l+1}^j \oplus H_{k_{j+1}}(t_{l+1}) \in U_{j+1}$ ou s'il existe $i \leq l$ tel que $x_{l+1}^j \oplus H_{k_{j+1}}(t_{l+1}) = x_i^j \oplus H_{k_{j+1}}(t_i)$, alors on choisit $P'_{j+1}(z_{l+1}^j \oplus H_{k_{j+1}}(t_{l+1}))$ uniformément aléatoirement dans $\{0,1\}^n \setminus (V_{j+1} \cup \{P'_{j+1}(z_i^j \oplus H_{k_{j+1}}(t_i)), i \leq l\})$ et il est impossible de coupler z_{l+1}^{j+1} et x_{l+1}^{j+1} à ce tour ;

   (b) sinon on définit $P'_{j+1}(z_{l+1}^j \oplus H_{k_{j+1}}(t_{l+1})) = P_{j+1}(x_{l+1}^j \oplus H_{k_{j+1}}(t_{l+1}))$, ainsi $z_{l+1}^{j+1} = x_{l+1}^{j+1}$.

La Propriété (ii) provient directement de ces règles et du fait que les clés et les *tweaks* soient identiques pour les deux schémas.

Vérification du fait que $(\mathbf{P}', \mathbf{k}')$ est uniformément aléatoire. Puisqu'on pose $\mathbf{k}' = \mathbf{k}$ et comme $\mathbf{k}$ est uniformément aléatoire, il en est de même pour $\mathbf{k}'$. Au cours du couplage des l premières requêtes, on pose $P'_j(x_i^{j-1} \oplus H_{k_j}(t_i)) = P_j(x_i^{j-1} \oplus H_{k_j}(t_i))$ pour chaque $i \leq l$ et $1 \leq j \leq r$; $P_j(x_i^{j-1} \oplus H_{k_j}(t_i))$ est uniformément aléatoire parmi les valeurs possibles donc il en est de même pour $P'_j(x_i^{j-1} \oplus H_{k_j}(t_i))$. La règle (1) dit que s'il y a collision avec une entrée précédente de P'_j, on ne peut pas choisir la

valeur de $P'_j(z_{l+1}^{j-1} \oplus H_{k_j}(t_i))$ donc ceci ne modifie pas la distribution de P'_j. Lorsque les conditions de la règle (2)(a) sont vérifiées, on a :

— pour un entier $i \leq l$:

$$\begin{cases} P_j(x_{l+1}^{j-1} \oplus H_{k_j}(t_{l+1})) = P_j(x_i^{j-1} \oplus H_{k_j}(t_i)) = P'_j(z_i^{j-1} \oplus H_{k_j}(t_i)) \\ z_{l+1}^{j-1} \oplus H_{k_j}(t_{l+1}) \neq z_i^{j-1} \oplus H_{k_j}(t_i), \end{cases}$$

— ou pour une paire $(u_j, v_j) \in \mathcal{Q}_{P_j}$:

$$\begin{cases} P_j(x_{l+1}^{j-1} \oplus H_{k_j}(t_{l+1})) = P_j(u_j) = P'_j(u_j) \\ z_{l+1}^{j-1} \oplus H_{k_j}(t_{l+1}) \neq u_j. \end{cases}$$

Dans les deux cas, on a que $P'_j(z_{l+1}^{j-1} \oplus H_{k_j}(t_{l+1})) \neq P_j(x_{l+1}^{j-1} \oplus H_{k_j}(t_{l+1}))$. Cela signifie que le couplage est impossible et on choisit $P'_j(z_{l+1}^{j-1} \oplus H_{k_j}(t_{l+1}))$ uniformément aléatoirement parmi les valeurs possibles pour garder P'_j uniformément distribuée. Finalement, lorsque les conditions de la règle (2)(b) sont vérifiées, il n'y a aucun problème pour coupler : $P_j(x_{l+1}^{j-1} \oplus H_{k_j}(t_{l+1}))$ et $P'_j(z_{l+1}^{j-1} \oplus H_{k_j}(t_{l+1}))$ sont tous les deux uniformément aléatoires parmi les valeurs possibles. Pour conclure, les permutations P'_j sont uniformément aléatoires et indépendantes comme voulu, de sorte que $(z_1^r, \ldots, z_{l+1}^r)$ est distribué selon ν'_l.

Probabilité de l'échec du Couplage. Il reste à majorer la probabilité que le couplage échoue, c'est-à-dire que

$$(z_1^r, \ldots, z_{l+1}^r) \neq (x_1^r, \ldots, x_{l+1}^r).$$

Pour chaque $j \in \{0, \ldots, r-1\}$, on note F_j l'évènement $z_{l+1}^j \oplus H_{k_{j+1}}(t_{l+1}) \in U_{j+1}$ ou $x_{l+1}^j \oplus H_{k_{j+1}}(t_{l+1}) \in U_{j+1}$ ou il existe $i \leq l$ tel que $z_{l+1}^j \oplus H_{k_{j+1}}(t_{l+1}) = z_i^j \oplus H_{k_{j+1}}(t_i)$ ou $x_{l+1}^j \oplus H_{k_{j+1}}(t_{l+1}) = x_i^j \oplus H_{k_{j+1}}(t_i)$. Il s'agit de l'évènement qui correspond à l'impossibilité de coupler au tour $j+1$. Alors on a :

$$\begin{aligned} \Pr[F_j] &\leq \sum_{i \leq l} \Pr[z_{l+1}^j \oplus H_{k_{j+1}}(t_{l+1}) = z_i^j \oplus H_{k_{j+1}}(t_i)] \\ &\quad + \sum_{i \leq l} \Pr[x_{l+1}^j \oplus H_{k_{j+1}}(t_{l+1}) = x_i^j \oplus H_{k_{j+1}}(t_i)] \\ &\quad + \sum_{u_{j+1} \in U_{j+1}} \Pr[z_{l+1}^j \oplus H_{k_{j+1}}(t_{l+1}) = u_{j+1}] \\ &\quad + \sum_{u_{j+1} \in U_{j+1}} \Pr[x_{l+1}^j \oplus H_{k_{j+1}}(t_{l+1}) = u_{j+1}] \\ &\leq \sum_{i \leq l} \Pr[H_{k_{j+1}}(t_i) \oplus H_{k_{j+1}}(t_{l+1}) = z_i^j \oplus z_{l+1}^j] \\ &\quad + \sum_{i \leq l} \Pr[H_{k_{j+1}}(t_i) \oplus H_{k_{j+1}}(t_{l+1}) = x_i^j \oplus x_{l+1}^j] \\ &\quad + \sum_{u_{j+1} \in U_{j+1}} \Pr[H_{k_{j+1}}(t_{l+1}) = z_{l+1}^j \oplus u_{j+1}] \\ &\quad + \sum_{u_{j+1} \in U_{j+1}} \Pr[H_{k_{j+1}}(t_{l+1}) = x_{l+1}^j \oplus u_{j+1}] \end{aligned}$$

$$\leq 2l\varepsilon + \frac{2q_p}{N},$$

où, pour la dernière inégalité, on a utilisé le fait que la famille de fonctions $\mathcal{H}$ est ε-AXU (on note que quand $t_{l+1} = t_j$, nécessairement $z^j_{l+1} \neq z^j_i$ et $x^j_{l+1} \neq x^j_i$ puisque $\mathcal{D}$ n'effectue jamais de requête inutile, de sorte que la probabilité est nulle) et l'uniformité de $\mathcal{H}$. Puisque les clés k_j sont indépendantes, on a :

$$\Pr\left[\bigcap_{i=0}^{r-1} F_i\right] \leq \left(2l\varepsilon + \frac{2q_p}{N}\right)^r. \tag{3.25}$$

En utilisant le lemme du couplage et le fait que $z^r_i = x^r_i$ pour tout $i \leq l$, on a :

$$\|\mu'_{l+1} - \mu'_l\| \leq \Pr\left[(z^r_1, \ldots, z^r_{l+1}) \neq (x^r_1, \ldots, x^r_{l+1})\right] \leq \Pr\left[z^r_{l+1} \neq x^r_{l+1}\right]. \tag{3.26}$$

Si on couple la dernière requête à un tour $j \leq r-1$, on sait que $z^{j'}_{l+1}$ et $x^{j'}_{l+1}$ restent égaux aux tours suivants, donc

$$\Pr\left[z^r_{l+1} \neq x^r_{l+1}\right] \leq \Pr\left[\bigcap_{i=0}^{r-1} F_i\right]. \tag{3.27}$$

En utilisant (3.25), (3.26) et (3.27), on a :

$$\|\mu'_{l+1} - \mu'_l\| \leq \left(2l\varepsilon + \frac{2q_p}{N}\right)^r \leq \left(2q_c\varepsilon + \frac{2q_p}{N}\right)^r. \tag{3.28}$$

Finalement, en utilisant (3.23) et (3.28), on obtient :

$$\|\mu_{\mathbf{t},\mathbf{x},\mathcal{Q}_\mathbf{P}} - \mu^*_\mathbf{t}\| \leq q_c \left(2q_c\varepsilon + \frac{2q_p}{N}\right)^r. \qquad \square$$

$\square$

Conclusion. En combinant les lemmes 20 et 21, on achève la preuve du théorème 6.

3.4.5 Du distingueur non adaptatif au distingueur adaptatif

Dans cette section, on considère le cas des distingueurs cca, et on prouve le résultat suivant.

Théorème 7. *Soient r un entier pair et $r' = r/2$. Soient q_c, q_p des entiers naturels, Alors*

$$\mathbf{Adv}^{\mathrm{cca}}_{\mathsf{TEM}[n,r,\mathbf{f}_\mathcal{H}]}(q_c, q_p) \leq \sqrt{2^{r'+4}\frac{q_c(N\varepsilon q_c + q_p)^{r'}}{N^{r'}}}.$$

Pour les entiers r impairs, on a $\mathbf{Adv}^{\mathrm{cca}}_{\mathsf{TEM}[n,r,\mathbf{f}_\mathcal{H}]} \leq \mathbf{Adv}^{\mathrm{cca}}_{\mathsf{TEM}[n,r-1,\mathbf{f}_\mathcal{H}]}$, de sorte que l'on peut utiliser le résultat précédent pour $r-1$ tours. En utilisant une famille de fonctions ε-AXU pour laquelle $\varepsilon \simeq 2^{-n}$, on voit que le schéma d'Even-Mansour

paramétrable avec un nombre pair r de tours est résistant aux attaques cca jusqu'à environ $2^{\frac{rn}{r+2}}$ requêtes de l'adversaire.

Afin de démontrer le théorème 7, nous allons utiliser le lemme 1 dans le cas particulier où toutes les transcriptions sont bonnes ($\varepsilon_2 = 0$). Dans ce but, nous allons dériver une borne appropriée ε_1 en doublant le nombre de tours de la construction et en utilisant le lemme 19. Notons que, dans cette sous-section, contrairement à la section 1.5.6, nous considérons seulement les transcriptions atteignables des requêtes $\tau' = (\mathcal{Q}_C, \mathcal{Q}_\mathbf{P})$ sans divulguer la clé $\mathbf{k}$ à l'adversaire à la fin de l'attaque. Nous noterons toujours T_{re} et T_{id} les variables aléatoires distribuées selon la distribution de probabilité de la transcription des requêtes τ' induite par le monde réel (respectivement idéal).

Lemme 22. *Soient r un entier naturel pair et $r' = r/2$. Soient q_c, q_p deux entiers naturels. On note :*

$$\alpha = 2^{r'} \frac{q_c(N\varepsilon q_c + q_p)^{r'}}{N^{r'}}.$$

Alors, pour toute transcription atteignable des requêtes τ', on a

$$\frac{\Pr[T_{\text{re}} = \tau']}{\Pr[T_{\text{id}} = \tau']} \geq 1 - 4\sqrt{\alpha}.$$

Démonstration. Fixons une transcription atteignable quelconque des requêtes $\tau' = (\mathcal{Q}_C, \mathcal{Q}_{P_1}, \ldots, \mathcal{Q}_{P_r})$. Comme précédemment, on désigne par $\mathbf{P} \vdash \mathcal{Q}_\mathbf{P}$ l'évènement $\wedge_{i=1}^{r}(P_i \vdash \mathcal{Q}_{P_i})$. Notons $\mathbf{P} = (P_1, \ldots, P_r)$, $\mathbf{P}'_1 = (P_1, \ldots, P_{r'})$ et $\mathbf{P}'_2 = (P_{r'+1}, \ldots, P_r)$. De même, $\mathbf{k}'_1 = (k_1, \ldots, k_{r'})$ et $\mathbf{k}'_2 = (k_{r'+1}, \ldots, k_r)$. On définit les évènements :

$$\mathbf{P}'_1 \vdash \mathcal{Q}_{\mathbf{P}'_1} = \wedge_{i=1}^{r'}(P_i \vdash \mathcal{Q}_{P_i}),$$
$$\mathbf{P}'_2 \vdash \mathcal{Q}_{\mathbf{P}'_2} = \wedge_{i=r'+1}^{r}(P_i \vdash \mathcal{Q}_{P_i}).$$

Soient également $\mathcal{Q}_C = ((t_1, x_1, y_1), \ldots, (t_{q_c}, x_{q_c}, y_{q_c}))$, $\mathbf{t} = (t_1, \ldots, t_{q_c})$, $\mathbf{x} = (x_1, \ldots, x_{q_c})$ et $\mathbf{y} = (y_1, \ldots, y_{q_c})$. Alors, pour chaque $i = 1, \ldots, q_c$,

$$\mathsf{TEM}_\mathbf{k}^\mathbf{P}(t_i, x_i) = \mathsf{TEM}_{\mathbf{k}'_2}^{\mathbf{P}'_2}\Big(t_i, \mathsf{TEM}_{\mathbf{k}'_1}^{\mathbf{P}'_1}(t_i, x_i)\Big).$$

Nous allons appliquer le lemme 19 indépendamment à chaque moitié du schéma $\mathsf{TEM}_{\mathbf{k}'_1}^{\mathbf{P}'_1}$ et $\mathsf{TEM}_{\mathbf{k}'_2}^{\mathbf{P}'_2}$. Considérons la première moitié $\mathsf{TEM}_{\mathbf{k}'_1}^{\mathbf{P}'_1}$. D'après le lemme 21, on a $\|\mu^1_{\mathbf{t},\mathbf{x},\mathcal{Q}_{\mathbf{P}'_1}} - \mu^*_\mathbf{t}\| \leq \alpha$, où $\mu^1_{\mathbf{t},\mathbf{x},\mathcal{Q}_{\mathbf{P}'_1}}$ est la distribution de $\mathsf{TEM}_{\mathbf{k}'_1}^{\mathbf{P}'_1}(\mathbf{t}, \mathbf{x})$ conditionnée par $\mathbf{P}'_1 \vdash \mathcal{Q}_{\mathbf{P}'_1}$. Ainsi le lemme 19 assure qu'il existe un sous-ensemble $S_\mathbf{x} \subset \Omega_\mathbf{t}$ de taille au moins $(1 - \sqrt{\alpha})|\Omega_\mathbf{t}|$ tel que, pour tout $\mathbf{z} \in S_\mathbf{x}$:

$$\mu^1_{\mathbf{t},\mathbf{x},\mathcal{Q}_{\mathbf{P}'_1}}(\mathbf{z}) \geq \frac{1 - \sqrt{\alpha}}{|\Omega_\mathbf{t}|}. \tag{3.29}$$

En appliquant un raisonnement similaire à la distribution $\mu^2_{\mathbf{t},\mathbf{y},\mathcal{Q}_{\mathbf{P}'_2}}$ de $(\mathsf{TEM}_{\mathbf{k}'_2}^{\mathbf{P}'_2})^{-1}(\mathbf{t}, \mathbf{y})$ conditionnée par $\mathbf{P}'_2 \vdash \mathcal{Q}_{\mathbf{P}'_2}$, on voit qu'il existe un sous-ensemble $S_\mathbf{y} \subset \Omega_\mathbf{t}$ de taille au moins $(1 - \sqrt{\alpha})|\Omega_\mathbf{t}|$ tel que, pour tout $\mathbf{z} \in S_\mathbf{y}$:

$$\mu^2_{\mathbf{t},\mathbf{y},\mathcal{Q}_{\mathbf{P}'_2}}(\mathbf{z}) \geq \frac{1 - \sqrt{\alpha}}{|\Omega_\mathbf{t}|}. \tag{3.30}$$

Notons que $|S_{\mathbf{x}} \cap S_{\mathbf{y}}| \geq (1 - 2\sqrt{\alpha})|\Omega_{\mathbf{t}}|$. Puisque les permutations $P_1, \ldots, P_r$ et les clés $k_1, \ldots, k_r$ sont uniformément aléatoires et indépendantes, on a

$$\begin{aligned}
\Pr[T_{\mathrm{re}} = \tau'] &\geq \sum_{\mathbf{z} \in S_{\mathbf{x}} \cap S_{\mathbf{y}}} \Pr\left[\mathsf{TEM}_{\mathbf{k}'_1}^{\mathbf{P}'_1}(\mathbf{t}, \mathbf{x}) = \mathbf{z} \wedge \mathbf{P}'_1 \vdash \mathcal{Q}_{\mathbf{P}'_1}\right] \\
&\qquad \times \Pr\left[(\mathsf{TEM}_{\mathbf{k}'_2}^{\mathbf{P}'_2})^{-1}(\mathbf{t}, \mathbf{y}) = \mathbf{z} \wedge \mathbf{P}'_2 \vdash \mathcal{Q}_{\mathbf{P}'_2}\right] \\
&\geq \sum_{\mathbf{z} \in S_{\mathbf{x}} \cap S_{\mathbf{y}}} \mu^1_{\mathbf{t},\mathbf{x},\mathcal{Q}_{\mathbf{P}'_1}}(\mathbf{z}) \mu^2_{\mathbf{t},\mathbf{y},\mathcal{Q}_{\mathbf{P}'_2}}(\mathbf{z}) \Pr[\mathbf{P}'_1 \vdash \mathcal{Q}_{\mathbf{P}'_1}] \Pr[\mathbf{P}'_2 \vdash \mathcal{Q}_{\mathbf{P}'_2}] \\
&\geq \Pr[\mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}}] \sum_{\mathbf{z} \in S_{\mathbf{x}} \cap S_{\mathbf{y}}} \mu^1_{\mathbf{t},\mathbf{x},\mathcal{Q}_{\mathbf{P}'_1}}(\mathbf{z}) \mu^2_{\mathbf{t},\mathbf{y},\mathcal{Q}_{\mathbf{P}'_2}}(\mathbf{z}).
\end{aligned}$$

En utilisant (3.29) et (3.30), on obtient

$$\begin{aligned}
\Pr[T_{\mathrm{re}} = \tau'] &\geq \Pr[\mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}}] \sum_{\mathbf{z} \in S_{\mathbf{x}} \cap S_{\mathbf{y}}} \frac{(1 - \sqrt{\alpha})^2}{|\Omega_{\mathbf{t}}|^2} \\
&\geq \Pr[\mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}}] \frac{(1 - 2\sqrt{\alpha})|S_{\mathbf{x}} \cap S_{\mathbf{y}}|}{|\Omega_{\mathbf{t}}|^2} \\
&\geq \Pr[\mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}}] \frac{(1 - 2\sqrt{\alpha})(1 - 2\sqrt{\alpha})}{|\Omega_{\mathbf{t}}|} \\
&\geq \Pr[\mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}}] \frac{1 - 4\sqrt{\alpha}}{|\Omega_{\mathbf{t}}|}. \qquad (3.31)
\end{aligned}$$

Considérons ensuite $\Pr[T_{\mathrm{id}} = \tau']$. Notons que, dans le monde idéal, le r-uplet de permutations est indépendant de la permutation paramétrable. Ainsi, on a

$$\Pr[T_{\mathrm{id}} = \tau'] = \Pr[\mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}}] \Pr[\widetilde{P} \vdash \mathcal{Q}_C].$$

Notons m le nombre de *tweaks* différents. Soit q_i le nombre de requêtes utilisant le i-ème *tweak*, en ordonnant les *tweaks* de façon arbitraire. Alors $|\Omega_{\mathbf{t}}| = \prod_{i=1}^{m} (N)_{q_i}$. Finalement, $\widetilde{P} \vdash \mathcal{Q}_C$ est équivalent à q_i contraintes sur la permutation associée avec le i-ème *tweak*, $i = 1, \ldots, m$. Alors

$$\Pr[\widetilde{P} \vdash \mathcal{Q}_C] = \prod_{i=1}^{m} \frac{1}{(N)_{q_i}} = \frac{1}{|\Omega_{\mathbf{t}}|}$$

et

$$\Pr[T_{\mathrm{id}} = \tau'] = \Pr[\mathbf{P} \vdash \mathcal{Q}_{\mathbf{P}}] \frac{1}{|\Omega_{\mathbf{t}}|}. \qquad (3.32)$$

Par conséquent, en combinant (3.31) et (3.32), on arrive à

$$\frac{\Pr[T_{\mathrm{re}} = \tau']}{\Pr[T_{\mathrm{id}} = \tau']} \geq 1 - 4\sqrt{\alpha}.$$

□

Conclusion. Combiner les lemmes 1 et 22 prouve le théorème 7.

Chapitre 4

Un mixage linéaire de la clé et du *tweak*

4.1 Introduction

Le chapitre précédent était concentré sur la construction d'un schéma d'Even-Mansour paramétrable mixant non-linéairement le *tweak* et la clé. Cette construction repose sur une famille ε-AXU et uniforme de fonctions. Bien que de telles familles de fonctions existent, il est possible que l'évaluation d'une telle fonction soit plus coûteuse que l'évaluation d'une permutation de tour. Par exemple, on peut noter que les processeurs récents de type x86 disposent de jeux d'instructions spécialisés dans l'évaluation de l'algorithme AES. On peut alors se demander, dans un souci d'efficacité et de simplicité, s'il est possible d'obtenir un schéma d'Even-Mansour paramétrable qui soit sûr, et qui mixe linéairement le *tweak* et la clé, éventuellement au prix d'un nombre de tours plus élevé.

Pour le reste du chapitre, fixons un entier naturel n et notons $N = 2^n$. Nous allons nous limiter au cas où l'espace des clés $\mathcal{K}$ est de la forme $\{0,1\}^\kappa$ pour un entier κ, et l'espace de tweak $\mathcal{T}$ est de la forme $\{0,1\}^\theta$ pour un entier θ.

Dans un premier temps, nous allons montrer que les constructions à un et deux tours ne peuvent fournir aucune sécurité. Nous allons ensuite exhiber deux constructions, à trois et quatre tours, qui seront résistantes aux attaques cca tant que le nombre de requêtes de l'adversaire est petit devant $2^{n/2}$, respectivement $2^{2n/3}$, où n est la taille du bloc. Ces travaux ont fait l'objet de deux publications aux conférences EUROCRYPT 2015 [CS15b] et ASIACRYPT 2015 [CS15a].

4.2 Trois tours sont nécessaires

4.2.1 Une attaque simple pour un tour

Commençons par une attaque très simple pour la construction à un tour. Étant donné une permutation P sur $\{0,1\}^n$ et deux applications linéaires $\gamma_0, \gamma_1 : \{0,1\}^\kappa \times \{0,1\}^\theta \to \{0,1\}^n$, considérons le schéma $\mathsf{TEM}[n, 1, (\gamma_0, \gamma_1)]$ qui, à partir d'une clé $k \in \{0,1\}^\kappa$, d'un tweak $t \in \{0,1\}^\theta$ et d'un message clair $x \in \{0,1\}^n$, calcule le

message chiffré défini par

$$\mathsf{TEM}^P(k,t,x) = \gamma_1(k,t) \oplus P(\gamma_0(k,t) \oplus x).$$

Étudions le distingueur qui ne soumet que deux requêtes à l'oracle de chiffrement $(0,x)$ et $(t, x \oplus \gamma_0(0,t))$, où $t \neq 0$, et reçoit comme réponse y et y' respectivement, puis vérifie si $y' = y \oplus \gamma_1(0,t)$. À cause de la linéarité de γ_0 et γ_1, cette égalité est vérifiée avec probabilité 1 dans le monde réel, mais seulement avec probabilité $1/N$ dans le monde idéal. L'avantage de cet adversaire est donc très proche de 1 : un tour de notre construction n'est donc pas suffisant pour offrir des garanties de sécurité prouvable. Nous allons également montrer qu'il existe un distingueur similaire pour deux tours de la construction TEM, lorsque la fonction de dérivation de clé est linéaire.

4.2.2 Une attaque pour deux tours

Nous exhibons maintenant une attaque plus sophistiquée capable de distinguer la construction TEM à deux tours (et, de nouveau, des fonctions de dérivation de clé qui sont linéaires en la clé et le tweak). Cette attaque ne nécessite aucune requête aux permutations internes, et ne requiert que quatre requêtes à l'oracle de construction. Elle peut être vue comme une attaque boomerang à clé reliée très efficace [BDK05]. Formellement, nous allons démontrer le théorème suivant.

Théorème 8. *Soit $\gamma = (\gamma_0, \gamma_1, \gamma_2)$ un triplet de fonctions linéaires de dérivation de clé. Alors*

$$\mathbf{Adv}^{\mathrm{cca}}_{\mathsf{TEM}[n,2,\gamma]}(4,0) \geq 1 - \frac{1}{N}.$$

Démonstration. Notons de façon générique $(\widetilde{P}, (P_1, P_2))$ les oracles auxquels l'adversaire a accès. Considérons le distingueur suivant (voir la figure 4.1 pour un diagramme représentant l'attaque) :

(1) choisir des valeurs arbitraires $x_1 \in \{0,1\}^n, t_1 \in \{0,1\}^\theta$, puis demander $y_1 := \widetilde{P}(t_1, x_1)$;

(2) choisir une valeur arbitraire $t_2 \in \{0,1\}^\theta \setminus \{t_1\}$, calculer $x_2 := x_1 \oplus \gamma_0(0, t_2 \oplus t_1)$, puis demander $y_2 := \widetilde{P}(t_2, x_2)$;

(3) choisir une valeur arbitraire $t_3 \in \{0,1\}^\theta \setminus \{t_1, t_2\}$, calculer $y_3 := y_1 \oplus \gamma_2(0, t_1 \oplus t_3)$, puis demander $x_3 := \widetilde{P}^{-1}(t_3, y_3)$;

(4) calculer $t_4 := t_3 \oplus t_2 \oplus t_1$ et $y_4 := y_2 \oplus \gamma_2(0, t_2 \oplus t_4)$, puis demander $x_4 := \widetilde{P}^{-1}(t_4, y_4)$;

(5) si $x_4 = x_3 \oplus \gamma_0(0, t_3 \oplus t_4)$, renvoyer 1, sinon renvoyer 0.

Quand le distingueur interagit avec le monde idéal $(\widetilde{P}, (P_1, P_2))$, où $\widetilde{P}$ est une permutation paramétrable indépendante de P_1 et P_2, la valeur x_4 est uniformément aléatoire et indépendante de x_3, t_3, et t_4 (en effet les tweaks t_i pour $i = 1, 2, 3, 4$ sont deux à deux distincts, donc y_4 est la première requête à la permutation aléatoire correspondant au tweak t_4). Ainsi, la probabilité que le distingueur renvoie 1 dans le cas idéal vaut 2^{-n}.

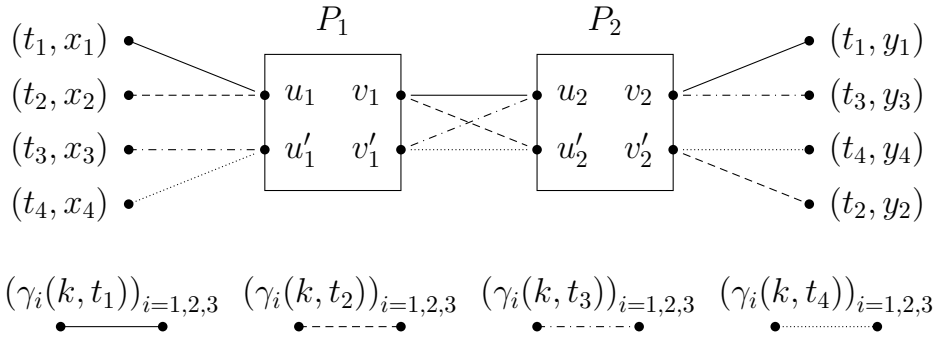

FIGURE 4.1 – Une attaque contre la construction d'Even-Mansour paramétrable à deux tours et des fonctions linéaires de dérivation de clé.

Montrons à présent que, lorsque le distingueur interagit avec le monde réel, c'est-à-dire avec $(\mathsf{TEM}_k^{P_1,P_2}, (P_1, P_2))$, il renvoie toujours 1, indépendamment de k, P_1, et P_2. Notons que, par définition, $x_2 = x_1 \oplus \gamma_0(0, t_2 \oplus t_1)$, on appelle u_1 la valeur commune

$$u_1 \stackrel{\text{def}}{=} x_1 \oplus \gamma_0(k, t_1) = x_2 \oplus \gamma_0(k, t_2),$$

et $v_1 = P_1(u_1)$. On note également

$$\begin{aligned} u_2 &= v_1 \oplus \gamma_1(k, t_1) && (4.1)\\ v_2 &= P_2(u_2) \\ u'_2 &= v_1 \oplus \gamma_1(k, t_2) && (4.2)\\ v'_2 &= P_2(u'_2). \end{aligned}$$

Ainsi, on a

$$\begin{aligned} y_1 &= v_2 \oplus \gamma_2(k, t_1) && (4.3)\\ y_2 &= v'_2 \oplus \gamma_2(k, t_2). && (4.4) \end{aligned}$$

Puisque $y_3 = y_1 \oplus \gamma_2(0, t_1 \oplus t_3)$, on peut voir, en utilisant (4.3) et la linéarité de γ_2, que

$$y_3 \oplus \gamma_2(k, t_3) = y_1 \oplus \gamma_2(k, t_1) = v_2.$$

Définissons

$$\begin{aligned} v'_1 &= u_2 \oplus \gamma_1(k, t_3) && (4.5)\\ u'_1 &= P_1^{-1}(v'_1). \end{aligned}$$

Ceci entraîne que

$$x_3 = u'_1 \oplus \gamma_0(k, t_3). \qquad (4.6)$$

Puisque $y_4 = y_2 \oplus \gamma_2(0, t_2 \oplus t_4)$, on voit d'après (4.4) et grâce à la linéarité de γ_2 que

$$y_4 \oplus \gamma_2(k, t_4) = y_2 \oplus \gamma_2(k, t_2) = v'_2.$$

De plus, comme $t_4 = t_3 \oplus t_2 \oplus t_1$, on a

$$\begin{aligned} u'_2 \oplus \gamma_1(k, t_4) &= u'_2 \oplus \gamma_1(k, t_2) \oplus \gamma_1(0, t_1 \oplus t_3) \\ &= v_1 \oplus \gamma_1(k, t_1) \oplus \gamma_1(k, t_3) && \text{d'après (4.2)} \\ &= u_2 \oplus \gamma_1(k, t_3) && \text{d'après (4.1)} \\ &= v'_1 && \text{d'après (4.5).} \end{aligned}$$

Cette égalité entraîne finalement que, d'après (4.6),

$$x_4 = u'_1 \oplus \gamma_0(k, t_4) = x_3 \oplus \gamma_0(0, t_3 \oplus t_4),$$

ce qui achève cette preuve. □

4.3 Une construction à 3 tours

4.3.1 Description de la construction

Les constructions à 1 et 2 tours pouvant être distinguées d'une permutation paramétrable idéale, nous allons maintenant considérer la construction TEM à trois tours munie des fonctions de dérivation de clé triviales : nous fixons $\mathcal{K} = \mathcal{T} = \{0,1\}^n$ et, pour tout $i = 0, \ldots, 3$ et tout $(k, t) \in \{0,1\}^n \times \{0,1\}^n$, on a $f_i(k, t) = k \oplus t$. Étant donné trois permutations P_1, P_2, P_3 de $\{0,1\}^n$, on note $\mathsf{TEM}[n,3,\mathbf{f}]^{P_1,P_2,P_3}$ le schéma d'Even-Mansour paramétrable TEM muni du quadruplet de fonctions de dérivation de clés triviales qui, à une clé $k \in \{0,1\}^n$, un *tweak* $t \in \{0,1\}^n$ et à un texte clair $x \in \{0,1\}^n$ associe le chiffré défini par

$$\mathsf{TEM}_k^{P_1,P_2,P_3}(t, x) = k \oplus t \oplus P_3(k \oplus t \oplus P_2(k \oplus t \oplus P_1(k \oplus t \oplus x))).$$

On prouve le résultat suivant [CS15b], qui a été découvert indépendamment par Farshim et Procter [FP15].

Théorème 9. *Soient q_e, q_p deux entiers naturels. On a*

$$\mathbf{Adv}^{\mathrm{cca}}_{\mathsf{TEM}[n,3,\mathbf{f}]}(q_c, q_p) \leq \frac{6q_cq_p}{N} + \frac{4q_c^2}{N}.$$

Démonstration. La preuve découle du lemme 1 combiné aux lemmes 23 et 24 démontrés ci-dessous. □

4.3.2 Description des mauvaises transcriptions

Définition 10. *Soit $\tau = (\mathcal{Q}_C, \mathcal{Q}_{P_1}, \mathcal{Q}_{P_2}, \mathcal{Q}_{P_3}, k)$ une transcription atteignable. On dit que τ est mauvaise si*

$$k \in \mathsf{BadK} = \bigcup_{1 \leq i \leq 2} \mathsf{BadK}_i$$

où :

$$k \in \mathsf{BadK}_1 \Leftrightarrow \textit{il existe } (t, x, y) \in \mathcal{Q}_C \textit{ et } (u_1, v_1) \in \mathcal{Q}_{P_1} \textit{ tels que } k \oplus t = x \oplus u_1$$

$k \in \mathsf{BadK}_2 \Leftrightarrow$ *il existe* $(t, x, y) \in \mathcal{Q}_C$ *et* $(u_3, v_3) \in \mathcal{Q}_{P_3}$ *tels que* $k \oplus t = y \oplus v_3$.

Dans le cas contraire, τ *est une bonne transcription. Notons* Θ_{bad} *l'ensemble des mauvaises transcriptions, et* $\Theta_{\text{good}} = \Theta \setminus \Theta_{\text{bad}}$ *l'ensemble des bonnes transcriptions.*

En premier lieu, majorons la probabilité d'obtenir une mauvaise transcription dans le monde idéal.

Lemme 23. *On a*

$$\Pr[T_{\text{id}} \in \Theta_{\text{bad}}] \leq \frac{2q_c q_p}{N}.$$

Démonstration. Puisque nous sommes dans le monde idéal, la clé k est tirée uniformément aléatoirement après la dernière requête, indépendamment de la transcription des requêtes. Ainsi, nous devons simplement majorer le nombre de mauvaises valeurs possibles pour k pour toutes les transcriptions atteignable des requêtes $(\mathcal{Q}_C, \mathcal{Q}_{P_1}, \mathcal{Q}_{P_2}, \mathcal{Q}_{P_3})$. Pour ce faire, fixons une quelconque transcription atteignable des requêtes $(\mathcal{Q}_C, \mathcal{Q}_{P_1}, \mathcal{Q}_{P_2}, \mathcal{Q}_{P_3})$. Alors, pour toute requête $(t, x, y) \in \mathcal{Q}_C$ et toute requête $(u_1, v_1) \in \mathcal{Q}_{P_1}$, il existe une unique clé k telle que $k = x \oplus t \oplus u_1$. Ainsi, $|\mathsf{BadK}_1| \leq q_c q_p$. De même, $|\mathsf{BadK}_2| \leq q_c q_p$. Ainsi, pour $i = 1, 2$,

$$\Pr\left[k \leftarrow_{\$} \{0,1\}^n : k \in \mathsf{BadK}_i\right] \leq \frac{q_c q_p}{N}.$$

On en déduit le résultat par l'inégalité de Boole. □

4.3.3 Étude des bonnes transcriptions

Lemme 24. *Pour toute bonne transcription* $\tau \in \Theta_{\text{good}}$*, on a*

$$\frac{\Pr\left[T_{\text{re}} = \tau\right]}{\Pr\left[T_{\text{id}} = \tau\right]} \geq 1 - \frac{4q_c q_p}{N} - \frac{4q_c^2}{N}.$$

Démonstration. Si $\Theta_{\text{good}} = \emptyset$, il n'y a rien à prouver. Sinon, fixons une bonne transcription $\tau = (\mathcal{Q}_C, \mathcal{Q}_{P_1}, \mathcal{Q}_{P_2}, \mathcal{Q}_{P_3}, k)$. Soit m le nombre de différents tweaks t qui apparaissent dans $\mathcal{Q}_C$ et q_i le nombre de requêtes utilisant le i-ème tweak (en ordonnant les tweaks de façon arbitraire). Notons que $q_c = \sum_{i=1}^{m} q_i$. Dans le monde idéal, on a simplement

$$\begin{aligned}\Pr[T_{\text{id}} = \tau] &= \Pr[k' \leftarrow_{\$} \{0,1\}^n : k' = k] \times \Pr[P_i \leftarrow_{\$} \mathsf{Perm}(n) : P_i \vdash \mathcal{Q}_{P_i}, i = 1,2,3] \\ &\qquad \times \Pr[E \leftarrow_{\$} \mathsf{TBC}(n,n) : (E,k) \vdash \mathcal{Q}_C] \\ &= \frac{1}{N} \cdot \frac{1}{((N)_{q_p})^3} \cdot \frac{1}{\prod_{i=1}^{m} (N)_{q_i}}. \end{aligned} \tag{4.7}$$

Nous devons simplement minorer la probabilité suivante :

$$\Pr\left[T_{\text{re}} = \tau\right] = \frac{1}{N} \times \Pr\left[P_1, P_2, P_3 \leftarrow_{\$} \mathsf{Perm}(n) : (\mathsf{EM}^{P_1,P_2,P_3}, k) \vdash \mathcal{Q}_E \wedge P_i \vdash \mathcal{Q}_{P_i}, i = 1,2,3\right].$$

Soient

$$\begin{aligned} U_1 &= \{u_1 \in \{0,1\}^n : (u_1, v_1) \in \mathcal{Q}_{P_1}\}, \quad & V_1 &= \{v_1 \in \{0,1\}^n : (u_1, v_1) \in \mathcal{Q}_{P_1}\}, \\ U_2 &= \{u_2 \in \{0,1\}^n : (u_2, v_2) \in \mathcal{Q}_{P_2}\}, \quad & V_2 &= \{v_2 \in \{0,1\}^n : (u_2, v_2) \in \mathcal{Q}_{P_2}\}, \\ U_3 &= \{u_3 \in \{0,1\}^n : (u_3, v_3) \in \mathcal{Q}_{P_3}\}, \quad & V_3 &= \{v_3 \in \{0,1\}^n : (u_3, v_3) \in \mathcal{Q}_{P_3}\} \end{aligned}$$

les ensembles de définition et image de $\mathcal{Q}_{P_1}$, $\mathcal{Q}_{P_2}$, et $\mathcal{Q}_{P_3}$ respectivement. Pour $u'_1 \in \{0,1\}^n$, soient $X(u'_1) = \{(t,x,y) \in \mathcal{Q}_C \,:\, x \oplus k \oplus t = u'_1\}$, et $U'_1 = \{u'_1 \in \{0,1\}^n : X(u'_1) \neq \emptyset\}$. De même, pour $v'_3 \in \{0,1\}^n$, soient $Y(v'_3) = \{(t,x,y) \in \mathcal{Q}_C : y \oplus k \oplus t = v'_3\}$, et $V'_3 = \{v'_3 \in \{0,1\}^n : Y(v'_3) \neq \emptyset\}$. Notons que, par définition d'une bonne transcription, on a $U_1 \cap U'_1 = \emptyset$ et $V_3 \cap V'_3 = \emptyset$. Soient également $\alpha = |U'_1|$ et $\beta = |V'_3|$. Par souci de clarté, on note

$$\begin{aligned} U'_1 &= \{u'_{1,1}, \ldots, u'_{1,\alpha}\} \\ V'_3 &= \{v'_{3,1}, \ldots, v'_{3,\beta}\} \end{aligned}$$

en ordonnant les éléments de façon arbitraire. Notons que

$$q_e = \sum_{i=1}^{\alpha} |X(u'_{1,i})| = \sum_{i=1}^{\beta} |Y(v'_{3,i})|. \tag{4.8}$$

Il est alors suffisant pour prouver notre résultat de minorer le nombre d'uplets de valeurs possibles pour $(v'_{1,1}, \ldots, v'_{1,\alpha})$ et $(u'_{3,1}, \ldots, u'_{3,\beta})$ tels que, en imposant $P_1(u'_{1,i}) = v'_{1,i}$ pour $1 \leq i \leq \alpha$ et $P_3(u'_{3,j}) = v'_{3,j}$ pour $1 \leq j \leq \beta$, l'évènement $E_k^{P_1,P_2,P_3} \vdash \mathcal{Q}_C$ est équivalent à q_c «nouvelles» équations sur P_2 (c'est-à-dire, distinctes des équations imposées par $P_2 \vdash \mathcal{Q}_{P_2}$). Plus précisément, soit N_1 le nombre d'uplets de valeurs deux à deux distinctes $(v'_{1,1}, \ldots, v'_{1,\alpha})$ tels que, pour tout $i = 1, \ldots, \alpha$:

(i) $v'_{1,i} \neq v_1$ pour tout $v_1 \in V_1$,

(ii) $v'_{1,i} \neq k \oplus t \oplus u_2$ pour tout $(t,x,y) \in X(u'_{1,i})$ et tout $u_2 \in U_2$,

(iii) $v'_{1,i} \neq t \oplus v'_{1,j} \oplus t'$ pour tout $(t,x,y) \in X(u'_{1,i})$, $1 \leq j \leq i-1$, $(t',x',y') \in X(u'_{1,j})$.

Alors

$$\begin{aligned} N_1 &\geq \prod_{i=1}^{\alpha} \left(N - q_p - i + 1 - |X(u'_{1,i})|(q_p + \sum_{j=1}^{i-1} |X(u'_{1,j})|) \right) \\ &\geq \prod_{i=1}^{\alpha} \left(N - q_p - q_e - |X(u'_{1,i})|(q_p + q_e) \right) \qquad \text{par (4.8).} \end{aligned}$$

De même, soit N_3 le nombre d'uplets de valeurs deux à deux distinctes pour $(u'_{3,1}, \ldots, u'_{3,\beta})$ tels que, pour tout $i = 1, \ldots, \beta$:

(i') $u'_{3,i} \neq u_3$ pour tout $u_3 \in U_3$,

(ii') $u'_{3,i} \neq k \oplus t \oplus v_2$ pour tout $(t,x,y) \in Y(v'_{3,i})$, $v_2 \in V_2$,

(iii') $u'_{3,i} \neq t \oplus u'_{3,j} \oplus t'$ pour tout $(t,x,y) \in Y(v'_{3,i})$, $1 \leq j \leq i-1$, $(t',x',y') \in Y(v'_{3,j})$.

Alors

$$\begin{aligned} N_3 &\geq \prod_{i=1}^{\beta} \left(N - q_p - i + 1 - |Y(v'_{3,i})|(q_p + \sum_{j=1}^{i-1} |Y(v'_{3,j})|) \right) \\ &\geq \prod_{i=1}^{\beta} \left(N - q_p - q_e - |Y(v'_{3,i})|(q_p + q_e) \right) \qquad \text{par (4.8).} \end{aligned}$$

Pour chaque choix possible de $(v'_{1,1}, \ldots, v'_{1,\alpha})$ et $(u'_{3,1}, \ldots, u'_{3,\beta})$ satisfaisant ces conditions, P_1 sera fixé sur exactement $q_p + \alpha$ points, P_2 sur $q_p + q_c$ points et P_3 sur $q_p + \beta$ points. Plus précisément, supposons $N_1 \cdot N_3 > 0$ et fixons deux uplets quelconques de valeurs $(v'_{1,1}, \ldots, v'_{1,\alpha})$ et $(u'_{3,1}, \ldots, u'_{3,\beta})$ satisfaisant ces conditions, et soit Ev_1 l'évènement $P_1(u'_{1,i}) = v'_{1,i}$ pour $1 \leq i \leq \alpha$ et Ev_3 l'évènement $P_3(u'_{3,j}) = v'_{3,j}$ pour $1 \leq j \leq \beta$. Alors, d'après les conditions (i) et (i') on a

$$\begin{aligned} \Pr\left[\mathsf{Ev}_1 \wedge (P_1 \vdash \mathcal{Q}_{P_1})\right] &= \frac{1}{(N)_{q_p+\alpha}} \\ \Pr\left[\mathsf{Ev}_3 \wedge (P_3 \vdash \mathcal{Q}_{P_3})\right] &= \frac{1}{(N)_{q_p+\beta}}. \end{aligned}$$

Fixons maintenant P_1 et P_3 satisfaisant Ev_1 et Ev_3. Pour chaque $(t, x, y) \in \mathcal{Q}_C$, soient respectivement u'_2 et v'_2 les entrées et sorties de P_2 pour cette requête, c'est-à-dire que $u'_2 = v'_{1,i} \oplus k \oplus t$ pour i tel que $x \oplus k \oplus t = u'_{1,i}$, et $v'_2 = u'_{3,j} \oplus k \oplus t$ for j tel que $y \oplus k \oplus t = v'_{3,j}$. Alors, les q_c valeurs u'_2 n'appartiennent pas à U_2 d'après la condition (ii), et sont deux à deux distinctes d'après la condition (iii). De même les q_c valeurs v'_2 ne sont pas dans V_2 d'après la condition (ii'), et sont deux à deux distinctes d'après la condition (iii'). Il suit que

$$\begin{aligned} &\Pr\left[(\mathsf{TEM}^{P_1,P_2,P_3}, k) \vdash \mathcal{Q}_C \wedge (P_2 \vdash \mathcal{Q}_{P_2}) \middle| \mathsf{Ev}_1 \wedge (P_1 \vdash \mathcal{Q}_{P_1}) \wedge \mathsf{Ev}_3 \wedge (P_3 \vdash \mathcal{Q}_{P_3})\right] \\ &\quad = \frac{1}{(N)_{q_p+q_c}}. \end{aligned}$$

Ainsi, en sommant sur les (au moins) $N_1 \cdot N_3$ paires possibles d'uplets, on obtient

$$\Pr\left[T_{\mathrm{re}} = \tau\right] \geq \frac{N_1 \cdot N_3}{N \cdot (N)_{q_p+\alpha} \cdot (N)_{q_p+q_e} \cdot (N)_{q_p+\beta}}. \tag{4.9}$$

Cette dernière inégalité est également vraie dans le cas où $N_1 \cdot N_3 = 0$. En utilisant (4.7) et (4.9), on a

$$\begin{aligned} \frac{\Pr\left[T_{\mathrm{re}} = \tau\right]}{\Pr\left[T_{\mathrm{id}} = \tau\right]} &\geq \frac{N_1 \cdot N_3 \cdot N \cdot (N)_{q_p}^3 \prod_{i=1}^{m} (N)_{q_i}}{N \cdot (N)_{q_p+\alpha} \cdot (N)_{q_p+q_c} \cdot (N)_{q_p+\beta}} \\ &\geq \frac{N_1 \cdot N_3 \cdot \prod_{i=1}^{m} (N)_{q_i}}{(N - q_p)_{\alpha} \cdot (N - q_p)_{q_c} \cdot (N - q_p)_{\beta}} \\ &\geq \frac{N_1 \cdot N_3 \cdot (N)_{q_c}}{(N - q_p)_{\alpha} \cdot (N - q_p)_{q_c} \cdot (N - q_p)_{\beta}} \end{aligned}$$

$$\geq \frac{N_1 \cdot N_3}{N^{\alpha+\beta}}.$$

Finalement, on a, puisque $\alpha \leq q_c$,

$$\begin{aligned}
\frac{N_1}{N^\alpha} =& \frac{\prod_{i=1}^{\alpha} \left(N - q_p - q_c - |X(u'_{1,i})|(q_p + q_c)\right)}{N^\alpha} \\
\geq& 1 - \sum_{i=1}^{\alpha} \frac{q_p + q_c + |X(u'_{1,i})|(q_p + q_c)}{N} \\
\geq& 1 - \frac{q_c q_p}{N} - \frac{q_c^2}{N} - (q_p + q_c) \sum_{i=1}^{\alpha} \frac{|X(u'_{1,i})|}{N} \\
\geq& 1 - \frac{2q_c q_p}{N} - \frac{2q_c^2}{N} \qquad \text{par (4.8).}
\end{aligned}$$

La même borne inférieure peut être prouvée pour $\frac{N_3}{N^\beta}$. Ainsi

$$\begin{aligned}
\frac{\Pr\left[T_{\text{re}} = \tau\right]}{\Pr\left[T_{\text{id}} = \tau\right]} &\geq \left(1 - \frac{2q_c q_p}{N} - \frac{2q_c^2}{N}\right)^2 \\
&\geq 1 - \frac{4q_c q_p}{N} - \frac{4q_c^2}{N}.
\end{aligned}$$

□

4.3.4 De l'exactitude de notre borne

Remarquons que notre construction est en fait un cas particulier de construction générique d'un algorithme de chiffrement par blocs paramétrable à partir d'un algorithme de chiffrement par blocs standard. En effet, si E est un algorithme de chiffrement par blocs, on peut définir un algorithme de chiffrement par blocs paramétrable $\widetilde{E}$, qui, pour toute clé k, tout *tweak* t, associe à tout message clair x le chiffré

$$\widetilde{E}_k(t, x) = E_{k \oplus t}(x).$$

Lorsque l'on remplace l'algorithme E par le schéma d'Even-Mansour à trois tours et utilisant des clés de tours identiques, on obtient le schéma précédent. Or il existe une attaque contre cette construction générique qui retrouve la clé en environ $2^{n/2}$ requêtes. L'attaque fonctionne de la façon suivante : l'adversaire effectue simplement des requêtes à l'oracle de construction sur les entrées $(t_i, 0)$ pour q_c valeurs t_i distinctes, et calcule $E_{k_j}(0)$ pour q_p valeurs distinctes k_j, ce qui se transpose, dans le modèle de la permutation aléatoire, en q_p requêtes à chaque permutation interne. Si les t_i et les k_j sont choisis tels que l'ensemble des valeurs $t_i \oplus k_j$ recouvre $\{0,1\}^n$, alors il existera une paire (i, j) telle que (dans le monde réel) $k^* \oplus t_i = k_j$, où k^* est la clé réellement utilisée dans l'oracle de construction. Ceci peut être détecté en recherchant une collision entre les deux listes $(\widetilde{P}(t_i, 0))_{1 \leq i \leq q_e}$ et $(E_{k_j}(0))_{1 \leq j \leq q_p}$, qui suggère un candidat pour la clé réelle k^*. Ceci entraîne en particulier que, avec une clé de n bits, la borne de sécurité ne peut pas être améliorée en augmentant le nombre de tours de notre construction.

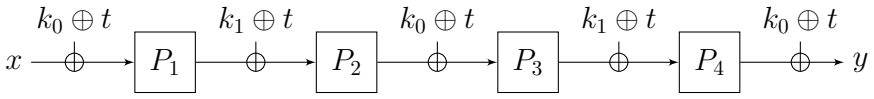

FIGURE 4.2 – La construction d'Even-Mansour paramétrable à 4 tours, avec une clé (k_0, k_1) de $2n$ bits et un tweak de n bits t.

4.4 Une construction à 4 tours

4.4.1 Description de la construction

Puisque nous ne pouvons pas espérer améliorer la sécurité de notre construction en augmentant simplement le nombre de tours, nous allons également doubler la taille de la clé maître. Les fonctions de dérivation de clé que nous allons utiliser seront alors définies par, pour $i = 0, ..., 4$, pour toute clé $(k_0, k_1) \in \{0,1\}^{2n}$ et tout *tweak* $t \in \{0,1\}^n$, $f_i((k_0, k_1), t) = k_{i \bmod 1} \oplus t$. Nous allons donc considérer la construction TEM à quatre tours $\mathsf{TEM}[n, 4, \mathbf{f}]$ utilisant des clés de $2n$ bits et des *tweaks* de n bits représentée sur la Figure 4.2. On prouve le résultat de sécurité suivant.

Théorème 10. *Soit* $\mathbf{f} = (f_0, \ldots, f_4)$ *le quadruplet de fonctions de* $\{0,1\}^{3n}$ *dans* $\{0,1\}^n$ *définies par* $f_i((k_0, k_1), t) = k_{i \bmod 2} \oplus t$. *Soient* q_c, q_p *deux entiers naturels tels que* $25n \leq q_c$ *et* $q_p + 3q_c + 1 \leq N/2$. *On a*

$$\mathbf{Adv}^{\mathrm{cca}}_{\mathsf{TEM}[n,4,\mathbf{f}]}(q_c, q_p) \leq \frac{44q_c^{3/2} + 38q_c\sqrt{q_p} + (30 + 5\sqrt{n})q_p\sqrt{q_c} + 4q_p^{3/2} + 2}{N}.$$

Ainsi, cette construction est résistante aux attaques cca tant que q_c et q_p sont petits devant $2^{2n/3}$, à un terme logarithmique en $N = 2^n$ près.

La preuve utilise la technique des coefficients H exposée dans la section 1.4. Dans la section 4.4.3, nous commençons par décrire l'ensemble des mauvaises transcriptions et majorons la probabilité d'obtenir une telle transcription dans le monde idéal. Ensuite,pour toute bonne transcription atteignable τ, nous prouvons dans la section 4.4.4 que le rapport entre la probabilité d'obtenir τ dans le monde réel et dans le monde imaginaire est suffisamment proche de 1.

4.4.2 Un lemme utile

Afin de majorer la probabilité d'obtenir une mauvaise transcription dans le monde idéal, nous aurons besoin d'une généralisation du «sum-capture theorem» de [CLL+14]. Les premiers résultats sur le «sum-capture problem» [Bab89, KPS13, Ste13] prouvent que, lorsque l'on choisit un sous-ensemble aléatoire A de $\mathbb{F}_2^n$ (ou plus généralement de n'importe quel groupe abélien) de taille q, la valeur

$$\mu(A) = \max_{\substack{U,V \subset \mathbb{F}_2^n \\ |U|=|V|=q}} |\{(a,u,v) \in A \times U \times V \,:\, a = u \oplus v\}|$$

est proche de son espérance q^3/N avec une probabilité très proche de 1 lorsque les trois ensembles sont choisis uniformément aléatoirement. Le problème considéré

dans [CLL+14] est du même type, mais se place dans le cas où l'ensemble A est construit à partir de l'interaction entre un distingueur et une permutation aléatoire P, plus précisément $A = \{x \oplus y \,:\, (x,y) \in \mathcal{Q}\}$, où $\mathcal{Q}$ est la transcription de l'interaction entre le distingueur et P. Nous allons étendre ce dernier résultat au cas où le distingueur interagit avec une famille de permutations aléatoires, autrement dit avec une permutation paramétrable aléatoire.

Notons $\mathsf{GL}(n)$ le groupe général linéaire de degré n sur $\mathbb{F}_2$, c'est-à-dire l'ensemble de tous les automorphismes d'espace vectoriel de $\mathbb{F}_2^n$.

Lemme 25. *Fixons un automorphisme $\Gamma \in \mathsf{GL}(n)$ et un ensemble non-vide $\mathcal{T}$. Soit $\widetilde{P}$ une permutation paramétrable uniformément aléatoire dans $\mathsf{TP}(\mathcal{T}, n)$, et soit $\mathcal{A}$ un algorithme probabiliste effectuant exactement q requêtes adaptatives et bidirectionnelles à $\widetilde{P}$. Soit $\widetilde{\mathcal{Q}} = ((t_1, x_1, y_1), \ldots, (t_q, x_q, y_q))$ la transcription de l'interaction de $\mathcal{A}$ avec $\widetilde{P}$. Pour tous sous-ensembles U et V de $\{0,1\}^n$, soit*

$$\mu(\widetilde{\mathcal{Q}}, U, V) = |\{((t,x,y),u,v) \in \widetilde{\mathcal{Q}} \times U \times V \,:\, x \oplus u = \Gamma(y \oplus v)\}|.$$

Alors, en supposant que $25n \leq q \leq N/2$, on a

$$\Pr_{\widetilde{P},\omega}\left[\exists U, V \subseteq \{0,1\}^n \,:\, \mu(\widetilde{\mathcal{Q}}, U, V) \geq \frac{q|U||V|}{N} + \frac{2q^2\sqrt{|U||V|}}{N} + 5\sqrt{nq|U||V|}\right] \leq \frac{2}{N},$$

où la probabilité est prise sur le choix aléatoire de $\widetilde{P}$ et les choix aléatoires ω de $\mathcal{A}$.

La preuve de ce lemme est une simple généralisation de celle qui est présente dans [CLL+14], nous l'incluons toutefois par souci d'exhaustivité. Celle-ci découle directement des lemmes 26 et 28 ci-dessous.

Rappels sur l'analyse de Fourier

Avant d'entamer la preuve des lemmes 26 et 28, il convient d'introduire de nouvelles notations et de rappeler quelques résultats classiques de l'analyse de Fourier dans le groupe abélien $(\mathbb{Z}/2\mathbb{Z})^n$ [CLL+14].

Étant donné un sous-ensemble $S \subset (\mathbb{Z}/2\mathbb{Z})^n$, on note $1_S \,:\, (\mathbb{Z}/2\mathbb{Z})^n \to \{0,1\}$ la fonction caractéristique de S, c'est-à-dire la fonction définie par, pour tout $x \in (\mathbb{Z}/2\mathbb{Z})^n$

$$1_S(x) = \begin{cases} 1 & \text{si } x \in S, \\ 0 & \text{sinon.} \end{cases}$$

Étant donné deux fonctions $f, g \,:\, (\mathbb{Z}/2\mathbb{Z})^n \to \mathbb{R}$, leur produit scalaire $\langle f, g\rangle$ est défini par

$$\langle f, g\rangle = \mathbb{E}[fg] = \frac{1}{N}\sum_{x \in (\mathbb{Z}/2\mathbb{Z})^n} f(x)g(x)$$

et, pour tout $x \in (\mathbb{Z}/2\mathbb{Z})^n$, on note

$$(f * g)(x) = \sum_{y \in (\mathbb{Z}/2\mathbb{Z})^n} f(y)g(x \oplus y)$$

le produit de convolution de f et g. Étant donné $\alpha \in (\mathbb{Z}/2\mathbb{Z})^n$, on note

$$\begin{aligned}\chi_\alpha \,:\, (\mathbb{Z}/2\mathbb{Z})^n &\to \{\pm 1\}\\ x &\mapsto (-1)^{\alpha\cdot x}\end{aligned}$$

le *caractère* associé à α. Le caractère associé à l'élément nul est appelé *caractère principal*, tandis que les autres sont les *caractères non-principaux*. On remarquera que l'ensemble des caractères forme un groupe pour la multiplication de fonctions et que l'on a, pour tout $\alpha, \beta \in (\mathbb{Z}/2\mathbb{Z})^n$, $\chi_\alpha \chi_\beta = \chi_{\alpha\oplus\beta}$.

Étant donné une fonction $f : (\mathbb{Z}/2\mathbb{Z})^n \to \mathbb{R}$ et $\alpha \in (\mathbb{Z}/2\mathbb{Z})^n$, le *coefficient de Fourier* de f correspondant à α est

$$\widehat{f}(\alpha) \stackrel{\text{def}}{=} \langle f, \chi_\alpha \rangle = \frac{1}{N} \sum_{x\in(\mathbb{Z}/2\mathbb{Z})^n} f(x)(-1)^{\alpha\cdot x}.$$

Le coefficient correspondant à l'élément nul est appelé le *coefficient de Fourier principal*, les autres sont appelés les *coefficients de Fourier non-principaux*. Notons que, pour tout ensemble $S \subset (\mathbb{Z}/2\mathbb{Z})^n$, on a

$$\widehat{1_S}(0) = \frac{|S|}{N}.$$

Nous utiliserons également les trois résultats classiques suivants, qui sont vrais pour toutes fonctions $f, g : (\mathbb{Z}/2\mathbb{Z})^n \to \mathbb{R}$, tout $\alpha \in (\mathbb{Z}/2\mathbb{Z})^n$ et tout $S \subset (\mathbb{Z}/2\mathbb{Z})^n$:

$$\sum_{x\in\{0,1\}^n} f(x)g(x) = N \sum_{\alpha\in\{0,1\}^n} \widehat{f}(\alpha)\widehat{g}(\alpha) \tag{4.10}$$

$$\widehat{(f * g)}(\alpha) = N\widehat{f}(\alpha)\widehat{g}(\alpha) \tag{4.11}$$

$$\sum_{\alpha\in\{0,1\}^n} |\widehat{1_S}(\alpha)|^2 = \frac{|S|}{N}. \tag{4.12}$$

Première étape : utilisation de l'inégalité de Cauchy-Schwartz

Dans ce premier lemme, nous allons majorer $\mu(\widetilde{\mathcal{Q}}, U, V)$ par une quantité qui ne dépend que de la transcription $\widetilde{\mathcal{Q}}$; plus précisément, cette majoration ne dépend que de l'amplitude maximale de la somme des coefficients non-principaux de Fourier des fonctions $\widehat{1_{\mathcal{Q}_t}}$ lorsque t parcourt $\mathcal{T}$. Cette démarche est à présent devenue classique, on peut la retrouver dans plusieurs articles qui étudient différentes variantes de ce problème [Bab89, Ste13, CLL+14].

Dans la suite, on note, pour tout $\alpha, \beta \in \{0,1\}^n$, $\alpha \neq 0$, $\beta \neq 0$,

$$\Phi_{\alpha,\beta}(\widetilde{\mathcal{Q}}) \stackrel{\text{def}}{=} N^2 \left|\sum_{t\in\mathcal{T}} \widehat{1_{\mathcal{Q}_t}}(\alpha,\beta)\right| = \left|\sum_{(t,x,y)\in\widetilde{\mathcal{Q}}} (-1)^{\alpha\cdot x\oplus\beta\cdot y}\right|$$

$$\Phi(\widetilde{\mathcal{Q}}) \stackrel{\text{def}}{=} \max_{\alpha\neq 0,\beta\neq 0} \Phi_{\alpha,\beta}(\widetilde{\mathcal{Q}})$$

où $\mathcal{Q}_t$ est l'ensemble des requêtes de $\mathcal{A}$ à la permutation $\widetilde{P}(t,\cdot)$ pour tout $t \in \mathcal{T}$.

Lemme 26. *Pour tous sous-ensembles U, V de $\{0,1\}^n$, on a*

$$\mu(\widetilde{\mathcal{Q}}, U, V) \leq \frac{q|U||V|}{N} + \Phi(\widetilde{\mathcal{Q}})\sqrt{|U||V|}.$$

Démonstration. Notons

$$\begin{aligned} W &= U \times V \\ K &= \{(\Gamma(k), k) \,:\, k \in \{0,1\}^n\}. \end{aligned}$$

Puisque, pour tout $t \in \mathcal{T}$, $((x,y),u,v) \in \mathcal{Q}_t \times U \times V$ satisfait $x \oplus u = \Gamma(y \oplus v)$ si et seulement si il existe $k \in \{0,1\}^n$ tel que

$$(x,y) \oplus (u,v) = (\Gamma(k), k),$$

on a

$$\begin{aligned} \mu(\widetilde{\mathcal{Q}}, U, V) &= \sum_{t\in\mathcal{T}} \sum_{(x,y)\in(\{0,1\}^n)^2} \sum_{(u,v)\in(\{0,1\}^n)^2} 1_{\mathcal{Q}_t}(x,y) 1_W(u,v) 1_K(x\oplus u, y \oplus v) \\ &= \sum_{t\in\mathcal{T}} \sum_{(x,y)\in(\{0,1\}^n)^2} 1_{\mathcal{Q}_t}(x,y) \sum_{(u,v)\in(\{0,1\}^n)^2} 1_W(u,v) 1_K(x\oplus u, y \oplus v) \\ &= \sum_{t\in\mathcal{T}} \sum_{(x,y)\in(\{0,1\}^n)^2} 1_{\mathcal{Q}_t}(x,y) (1_W * 1_K)(x,y) \\ &= N^2 \sum_{t\in\mathcal{T}} \sum_{(\alpha,\beta)\in(\{0,1\}^n)^2} \widehat{1_{\mathcal{Q}_t}}(x,y) \widehat{(1_W * 1_K)}(\alpha,\beta) \qquad \text{d'après (4.10)} \\ &= N^4 \sum_{t\in\mathcal{T}} \sum_{(\alpha,\beta)\in(\{0,1\}^n)^2} \widehat{1_{\mathcal{Q}_t}}(x,y) \widehat{1_W}(\alpha,\beta) \widehat{1_K}(\alpha,\beta) \qquad \text{d'après (4.11).} \end{aligned}$$

En séparant le coefficient de Fourier principal des coefficients non-principaux dans la dernière inégalité ci-dessus, on obtient

$$\begin{aligned} \mu(\widetilde{\mathcal{Q}}, U, V) &= N^4 \sum_{t\in\mathcal{T}} \frac{|\mathcal{Q}_t|}{N^2} \frac{|W|}{N^2} \frac{|K|}{N^2} + N^4 \sum_{t\in\mathcal{T}} \sum_{\substack{(\alpha,\beta) \\ \neq(0,0)}} \widehat{1_{\mathcal{Q}_t}}(x,y) \widehat{1_W}(\alpha,\beta) \widehat{1_K}(\alpha,\beta) \\ &= \frac{q|U||V|}{N} + N^4 \sum_{t\in\mathcal{T}} \sum_{(\alpha,\beta)\neq(0,0)} \widehat{1_{\mathcal{Q}_t}}(x,y) \widehat{1_W}(\alpha,\beta) \widehat{1_K}(\alpha,\beta). \qquad (4.13) \end{aligned}$$

De plus, on a

$$\begin{aligned} \widehat{1_W}(\alpha,\beta) &= \frac{1}{N^2} \sum_{(u,v)\in(\{0,1\}^n)^2} 1_W(u,v) (-1)^{\alpha\cdot u \oplus \beta \cdot v} \\ &= \frac{1}{N^2} \sum_{(u,v)\in(\{0,1\}^n)^2} 1_U(u) 1_V(v) (-1)^{\alpha\cdot u \oplus \beta \cdot v} \\ &= \frac{1}{N^2} \left(\sum_{u\in\{0,1\}^n} 1_U(u) (-1)^{\alpha\cdot u} \right) \left(\sum_{v\in\{0,1\}^n} 1_V(v) (-1)^{\beta\cdot v} \right) \\ &= \widehat{1_U}(\alpha) \widehat{1_V}(\beta), \end{aligned}$$

et

$$\begin{aligned}\widehat{1_K}(\alpha,\beta) &= \frac{1}{N^2}\sum_{(x,y)\in(\{0,1\}^n)^2} 1_K(x,y)(-1)^{\alpha\cdot x\oplus\beta\cdot y}\\ &= \frac{1}{N^2}\sum_{y\in\{0,1\}^n}(-1)^{\alpha\cdot\Gamma(y)\oplus\beta\cdot y}\\ &= \frac{1}{N^2}\sum_{y\in\{0,1\}^n}(-1)^{\Gamma^*(\alpha)\cdot y\oplus\beta\cdot y}\\ &= 0 \text{ if } \beta\neq\Gamma^*(\alpha)\\ &\quad \frac{1}{N} \text{ if } \beta=\Gamma^*(\alpha).\end{aligned}$$

Alors, en injectant ces deux observations dans (4.13), on obtient

$$\begin{aligned}\mu(\widetilde{\mathcal{Q}},U,V) &= \frac{q|U||V|}{N} + N^3\sum_{\alpha\neq 0}\widehat{1_U}(\alpha)\widehat{1_V}(\Gamma^*(\alpha))\sum_{t\in\mathcal{T}}\widehat{1_{\mathcal{Q}_t}}(\alpha,\Gamma^*(\alpha))\\ &\leq \frac{q|U||V|}{N} + N^3\sum_{\alpha\neq 0}\left|\widehat{1_U}(\alpha)\right|\cdot\left|\widehat{1_V}(\Gamma^*(\alpha))\right|\cdot\left|\sum_{t\in\mathcal{T}}\widehat{1_{\mathcal{Q}_t}}(\alpha,\Gamma^*(\alpha))\right|\\ &\leq \frac{q|U||V|}{N} + N\Phi(\widetilde{\mathcal{Q}})\sum_{\alpha\neq 0}\left|\widehat{1_U}(\alpha)\right|\cdot\left|\widehat{1_V}(\Gamma^*(\alpha))\right|,\end{aligned}$$

par définition de $\Phi(\widetilde{\mathcal{Q}})$. L'inégalité de Cauchy-Schwartz donne

$$\begin{aligned}\sum_{\alpha\neq 0}\left|\widehat{1_U}(\alpha)\right|\cdot\left|\widehat{1_V}(\Gamma^*(\alpha))\right| &\leq \sqrt{\sum_{\alpha\in\{0,1\}^n}|\widehat{1_U}(\alpha)|^2}\sqrt{\sum_{\alpha\in\{0,1\}^n}|\widehat{1_V}(\Gamma^*(\alpha))|^2}\\ &= \frac{1}{N}\sqrt{|U||V|},\end{aligned}$$

où, pour la dernière inégalité, on a utilisé l'égalité (4.12), de telle sorte qu'on obtient

$$\mu(\widetilde{\mathcal{Q}},U,V) \leq \frac{q|U||V|}{N} + \Phi(\widetilde{\mathcal{Q}})\sqrt{|U||V|}.$$

□

Deuxième étape : majoration de $\Phi(\widetilde{\mathcal{Q}})$

Il reste maintenant à trouver une majoration de $\Phi(\widetilde{\mathcal{Q}})$ qui soit vraie avec une forte probabilité sur le choix de la permutation paramétrable $\widetilde{P}$ et l'aléa de l'adversaire. Pour ce faire, nous allons utiliser le lemme suivant, tiré de [CLL+14], qui constitue une généralisation de la borne de Chernoff dans le cas de variables aléatoires qui ne sont pas nécessairement indépendantes, mais pour lesquelles cette dépendance reste modérée.

Lemme 27 ([CLL+14])**.** *Soit $0 \leq \epsilon < 1/2$, et soit $\mathbf{A} = (A_i)_{1\leq i\leq q}$ une suite de variables aléatoires à valeur dans $\{1,-1\}$. Supposons que, pour tout $1 \leq i \leq q$ et toute suite $(a_1,\ldots,a_{i-1}) \in \{1,-1\}^{i-1}$, on a*

$$\Pr\left[A_i = 1 \,|\, (A_1,\ldots,A_{i-1}) = (a_1,\ldots,a_{i-1})\right] \leq \frac{1}{2} + \epsilon.$$

Alors, pour tout $\delta \in [0,1]$, on a

$$\Pr\left[\sum_{i=1}^{q} A_i \geq q(2\epsilon+\delta)\right] \leq e^{-\frac{q\delta^2}{12}}.$$

Démonstration. Remarquons tout d'abord que, si $\varepsilon = 1/2$, alors le théorème est trivial. Nous allons donc supposer que $\varepsilon < 1/2$.

Soit $\mathbf{A} = (A_i)_{1\leq i\leq q}$ une suite de variables aléatoires indépendantes et identiquement distribuées et telles que, pour tout $i \in \{1,\ldots,q\}$,

$$\Pr[B_i = 1] = \frac{1}{2} + \varepsilon \qquad \text{et} \qquad \Pr[B_i = -1] = \frac{1}{2} - \varepsilon.$$

La première étape de la preuve de ce résultat est de montrer que, pour tout r, on a

$$\Pr\left[\sum_{i=1}^{q} A_i \geq r\right] \leq \Pr\left[\sum_{i=1}^{q} B_i \geq r\right].$$

Ceci se démontre par un argument similaire à ceux que l'on peut trouver dans une preuve utilisant la méthode du couplage. Notons μ_p la distribution de probabilité de paramètre $0 \leq p \leq 1$ telle que, si X est une variable aléatoire suivant la loi μ_p, alors

$$\Pr[X = x] = \begin{cases} p & \text{si } x = 1 \\ 1-p & \text{si } x = -1 \\ 0 & \text{sinon.} \end{cases}$$

Considérons l'échantillonnage suivant :

for $i = 1$ to q **do**
  $p \leftarrow \Pr\left[A_i = 1 \,|\, (A_1,\ldots,A_{i-1}) = (u_1,\ldots,u_{i-1})\right]$
  $u_i \leftarrow_{\$} \mu_p$
  if $u_i = 1$ **then**
    $v_i \leftarrow 1$
  else
    $p' \leftarrow \frac{1/2+\varepsilon-p}{1-p}$
    $v_i \leftarrow_{\$} \mu_{p'}$
  end if
end for
return $((u_1,\ldots,u_q),(v_1,\ldots,v_q))$

Alors, par définition, la loi de probabilité suivie par le q-uplet $(u_1,\ldots,u_q)$ est identique à celle de $\mathbf{A}$, tandis que le q-uplet $(v_1,\ldots,v_q)$ suit la même loi de probabilité que $\mathbf{B}$. En effet, pour tout $i = 1,\ldots,q$ et toute suite $(v_1,\ldots,v_{i-1}) \in \{1,-1\}^{i-1}$, on a

$$\Pr[v_i = 1 | (v_1,\ldots,v_{i-1})] = p + p'(1-p) = \frac{1}{2} + \varepsilon.$$

Notons que, par définition du processus d'échantillonnage, si $u_i = 1$, alors $v_i = 1$, ce qui entraîne que, pour tout $i = 1, \ldots, q$, on a $v_i \geq u_i$. Ainsi, on a

$$\sum_{i=1}^{q} u_i \geq r \implies \sum_{i=1}^{q} v_i \geq r$$

et par conséquent

$$\Pr\left[\sum_{i=1}^{q} A_i \geq r\right] \leq \Pr\left[\sum_{i=1}^{q} B_i \geq r\right]. \tag{4.14}$$

Fixons un réel $\delta \in [0,1]$, et soit $(B'_i)_{1\leq i\leq q}$ la suite de variables aléatoires définie par, pour $i = 1, \ldots, q$,

$$B'_i = \frac{1 + B_i}{2},$$

de sorte que B'_i suit une loi de Bernoulli de paramètre $\frac{1}{2} + \varepsilon$, c'est-à-dire que

$$\Pr[B'_i = 1] = \frac{1}{2} + \varepsilon \qquad \text{et} \qquad \Pr[B'_i = 0] = \frac{1}{2} - \varepsilon.$$

Soit m l'espérance de la somme des variables aléatoires B'_i, pour $i = 1, \ldots, q$. On a alors

$$m = \mathbb{E}\left[\sum_{i=1}^{q}\right] = q\left(\frac{1}{2} + \varepsilon\right).$$

La borne de Chernoff, spécialisée au cas de variables aléatoires de Bernoulli, assure que, pour tout réel $0 \leq \delta' \leq 1$, on a

$$\Pr\left[\sum_{i=1}^{q} B'_i \geq (1+\delta')m\right] \leq e^{\frac{-m\delta'^2}{3}}.$$

En choisissant

$$\delta' = \frac{q\delta}{2m} = \frac{\delta}{1+2\varepsilon}$$

dans l'inégalité précédente, et en remarquant que, si $\delta \in [0,1]$, alors $\delta' \in [0,1]$, on obtient

$$\begin{aligned}\Pr\left[\sum_{i=1}^{q} B_i \geq q(2\varepsilon+\delta)\right] &= \Pr\left[\sum_{i=1}^{q} B'_i \geq \left(1 + \frac{q\delta}{2m}\right)m\right] \\ &\leq e^{-\frac{q^2\delta^2}{12m}} \leq e^{\frac{q\delta^2}{12}}.\end{aligned}$$

Combiner cette dernière inégalité avec l'équation (4.14) achève cette preuve. □

Il est désormais possible de majorer $\Phi(\widetilde{\mathcal{Q}})$.

Lemme 28. *Supposons que $25n \leq q \leq N/2$. Fixons un adversaire $\mathcal{A}$ effectuant q requêtes pour une permutation paramétrable aléatoire $\widetilde{P}$ ayant pour espace de tweak $\mathcal{T}$. Soit $\widetilde{\mathcal{Q}}$ la transcription de l'interaction de $\mathcal{A}$ avec $\widetilde{P}$. Alors*

$$\Pr_{\widetilde{P},\omega}\left[\Phi(\widetilde{\mathcal{Q}}) \geq \frac{2q^2}{N} + 5\sqrt{nq}\right] \leq \frac{2}{N},$$

où la probabilité est prise sur le choix aléatoire de $\widetilde{P}$ et les choix aléatoires ω de $\mathcal{A}$.

Démonstration. Dans cette preuve, on écrit $\Pr[\cdot]$ pour $\Pr_{\widetilde{P},\omega}[\cdot]$. Fixons $\alpha, \beta \in \{0,1\}^n$, $\alpha \neq 0$ et $\beta \neq 0$. On note $\widetilde{\mathcal{Q}} = ((t_1, x_1, y_1), \ldots, (t_q, x_q, y_q))$ en suivant l'ordre naturel des requêtes de $\mathcal{A}$, et on définit la suite de variables aléatoires $(A_i)_{1 \leq i \leq q}$ où $A_i = (-1)^{\alpha \cdot x_i \oplus \beta \cdot y_i}$. Alors $\Phi_{\alpha,\beta}(\widetilde{\mathcal{Q}}) = |\sum_{i=1}^{q} A_i|$. Afin d'appliquer le lemme 27, nous allons prouver que, pour tout $1 \leq i \leq q$, et toute suite $(a_1, \ldots, a_{i-1}) \in \{1, -1\}^{i-1}$, on a

$$p_i \stackrel{\text{def}}{=} \Pr[A_i = 1 \mid (A_1, \ldots, A_{i-1}) = (a_1, \ldots, a_{i-1})] \leq \frac{1}{2} + \frac{q}{N}. \tag{4.15}$$

La $i^{\text{ème}}$ requête concerne la permutation $P(t_i, \cdot)$. On suppose qu'il s'agit d'une requête directe x_i. Remarquons que la réponse y_i est distribuée uniformément aléatoirement dans un ensemble de taille au moins $N - i + 1$. Observons également que, une fois que x_i est fixé, il y a exactement $N/2$ valeurs pour y_i telles que $A_i = (-1)^{\alpha \cdot x_i \oplus \beta \cdot y_i} = 1$ puisque $\beta \neq 0$. De même, si la $i^{\text{ème}}$ requête est une requête inverse y_i, alors la réponse x_i est distribuée uniformément aléatoirement dans un ensemble de taille au moins $N - i + 1$, et une fois que y_i est fixé, il y a exactement $N/2$ valeurs possibles pour x_i telles que $A_i = (-1)^{\alpha \cdot x_i \oplus \beta \cdot y_i} = 1$ car $\alpha \neq 0$. Ainsi on a que

$$p_i \leq \frac{N/2}{N - i + 1} \leq \frac{N}{2(N-q)} \leq \frac{1}{2} + \frac{q}{2(N-q)} \leq \frac{1}{2} + \frac{q}{N}.$$

On peut maintenant utiliser le lemme 27 avec $\epsilon = \frac{q}{N}$ et on obtient, pour tout $\delta \in [0,1]$,

$$\Pr\left[\sum_{i=1}^{q} A_i \geq \frac{2q^2}{N} + q\delta\right] \leq e^{-\frac{q\delta^2}{12}}.$$

Définissons $A'_i = -A_i$, et, en appliquant exactement le même raisonnement, on obtient

$$\Pr\left[\sum_{i=1}^{q} A_i \leq -\frac{2q^2}{N} + q\delta\right] \leq e^{-\frac{q\delta^2}{12}}.$$

Ainsi, par l'inégalité de Boole, on a

$$\Pr\left[\Phi_{\alpha,\beta}(\widetilde{\mathcal{Q}}) \geq \frac{2q^2}{N} + q\delta\right] \leq 2e^{-\frac{q\delta^2}{12}}.$$

Notons que cette inégalité est vraie pour tout $\alpha \neq 0$ et tout $\beta \neq 0$. Ainsi, en choisissant $\delta = \sqrt{(36 \ln N)/q}$ (qui, en supposant $q \geq 25n$, vérifie $\delta < 1$), et en utilisant le fait que $\sqrt{36 \ln 2} < 5$, on obtient finalement que, pour tout $\alpha \neq 0$ et tout $\beta \neq 0$,

$$\Pr\left[\Phi_{\alpha,\beta}(\widetilde{\mathcal{Q}}) \geq \frac{2q^2}{N} + \sqrt{5nq}\right] \leq \frac{2}{N^3}$$

et le résultat suit en sommant sur les valeurs possibles pour α et β. □

4.4.3 Description des mauvaises transcriptions

La première étape est de définir l'ensemble des mauvaises transcriptions. Soit $\tau = (\mathcal{Q}_C, \mathcal{Q}_{P_1}, \ldots, \mathcal{Q}_{P_4}, (k_0, k_1))$ une transcription atteignable, où $|\mathcal{Q}_C| = q_c$ et $|\mathcal{Q}_{P_i}| = q_p$ pour $i = 1, \ldots, 4$. Dans le reste de la section, on note, pour $i \in \{1, \ldots, 4\}$,

$$U_i = \{u_i \in \{0,1\}^n : (u_i, v_i) \in \mathcal{Q}_{P_i}\}$$

$$V_i = \{v_i \in \{0,1\}^n : (u_i, v_i) \in \mathcal{Q}_{P_i}\}$$

respectivement les ensembles de définition et des images de $\mathcal{Q}_{P_i}$. On définit également trois nouvelles quantités caractérisant une transcription :

$$\begin{aligned}
\alpha_1 &\stackrel{\text{def}}{=} |\{((t,x,y),u_1) \in \mathcal{Q}_C \times U_1 \,:\, x \oplus k_0 \oplus t = u_1\}| \\
\alpha_4 &\stackrel{\text{def}}{=} |\{((t,x,y),v_4) \in \mathcal{Q}_C \times V_4 \,:\, y \oplus k_0 \oplus t = v_4\}| \\
\alpha_{2,3} &\stackrel{\text{def}}{=} |\{((t,x,y),v_2,u_3) \in \mathcal{Q}_C \times V_2 \times U_3 \,:\, v_2 \oplus k_0 \oplus t = u_3\}|.
\end{aligned}$$

On introduit aussi deux quantités dépendant respectivement de $\mathcal{Q}_{P_2}$ et $\mathcal{Q}_{P_3}$:

$$\begin{aligned}
\nu_2 &\stackrel{\text{def}}{=} |\{((u_2,v_2),(u_2',v_2')) \in (\mathcal{Q}_{P_2})^2 \,:\, (u_2,v_2) \neq (u_2',v_2'),\, u_2 \oplus v_2 = u_2' \oplus v_2'\}| \\
\nu_3 &\stackrel{\text{def}}{=} |\{((u_3,v_3),(u_3',v_3')) \in (\mathcal{Q}_{P_3})^2 \,:\, (u_3,v_3) \neq (u_3',v_3'),\, u_3 \oplus v_3 = u_3' \oplus v_3'\}|.
\end{aligned}$$

Définition 11. *Une transcription τ est dite mauvaise si au moins l'une des conditions suivantes est satisfaite :*

(B-1) il existe $(t,x,y) \in \mathcal{Q}_C$, $(u_1,v_1) \in \mathcal{Q}_{P_1}$, et $(u_4,v_4) \in \mathcal{Q}_{P_4}$ tels que $k_0 \oplus t = x \oplus u_1 = v_4 \oplus y$;

(B-2) il existe $(t,x,y) \in \mathcal{Q}_C$, $(u_1,v_1) \in \mathcal{Q}_{P_1}$, et $(u_2,v_2) \in \mathcal{Q}_{P_2}$ tels que $k_0 \oplus t = x \oplus u_1$ et $k_1 \oplus t = v_1 \oplus u_2$;

(B-3) il existe $(t,x,y) \in \mathcal{Q}_C$, $(u_3,v_3) \in \mathcal{Q}_{P_3}$, et $(u_4,v_4) \in \mathcal{Q}_{P_4}$ tels que $k_1 \oplus t = v_3 \oplus u_4$ et $k_0 \oplus t = v_4 \oplus y$;

(B-4) $\alpha_1 \geq \sqrt{q_c}/2$;

(B-5) $\alpha_4 \geq \sqrt{q_c}/2$;

(B-6) $\alpha_{2,3} \geq q_p\sqrt{q_c}$;

(B-7) $\nu_2 \geq \sqrt{q_p}$;

(B-8) $\nu_3 \geq \sqrt{q_p}$.

Dans le cas contraire τ est dite bonne. On note Θ_{good}, respectivement Θ_{bad} l'ensemble des bonnes, respectivement des mauvaises transcriptions.

Les conditions (B-4) et (B-5) sont définies de la sorte pour nous permettre d'utiliser directement le lemme 17. Débutons par majorer la probabilité d'obtenir une mauvaise transcription dans le monde idéal.

Lemme 29. *Supposons que $25n \leq q_c \leq N/2$ et $q_p \leq N/2$. On a*

$$\Pr\left[T_{\text{id}} \in \Theta_{\text{bad}}\right] \leq \frac{2q_c^2 q_p + 3q_c q_p^2}{N^2} + \frac{(5+3\sqrt{n})\sqrt{q_c}q_p + 4q_p^{3/2} + 2}{N}.$$

Démonstration. On majore successivement les probabilités que chaque condition soit vraie dans le monde idéal. Notons Θ_i l'ensemble des transcriptions atteignables satisfaisant la condition (B-i). On rappelle que, dans le monde idéal, la clé (k_0,k_1) est choisie aléatoirement, indépendamment de la transcription des requêtes.

Condition (B-1). Soit BadK_1 l'ensemble des clés k_0 telles qu'il existe $(t, x, y) \in \mathcal{Q}_C$, $(u_1, v_1) \in \mathcal{Q}_{P_1}$, et $(u_4, v_4) \in \mathcal{Q}_{P_4}$ tels que $k_0 \oplus t = x \oplus u_1 = y \oplus v_4$. Notons que BadK_1 dépend uniquement de la transcription des requêtes, ainsi, pour toute constante C, on a, puisque k_0 est uniformément aléatoire,

$$\Pr\left[T_{\mathrm{id}} \in \Theta_1\right] \leq \Pr\left[\widetilde{P}_0 \leftarrow_{\$} \mathsf{TP}(\mathcal{T}, n), \mathbf{P} \leftarrow_{\$} (\mathsf{P}(n))^4 \,:\, |\mathsf{BadK}_1| > C\right] + \frac{C}{N}. \quad (4.16)$$

De plus, en posant

$$\mu(\mathcal{Q}_C, U_1, V_4) \stackrel{\text{def}}{=} |\{((t, x, y), u_1, v_4) \in \mathcal{Q}_C \times U_1 \times V_4 \,:\, x \oplus u_1 = y \oplus v_4)\}|,$$

alors on a clairement

$$|\mathsf{BadK}_1| \leq \mu(\mathcal{Q}_C, U_1, V_4).$$

Ainsi, on peut utiliser le lemme 25 afin de majorer $|\mathsf{BadK}_1|$ avec une probabilité écrasante (on considère $\mathcal{D}$ avec un accès aux permutations internes comme un algorithme probabiliste $\mathcal{A}$ interagissant avec la permutation paramétrable $\widetilde{P}_0$, aboutissant à la transcription $\mathcal{Q}_C$). Pour

$$C = \frac{q_c q_p^2}{N} + \frac{2 q_c^2 q_p}{N} + 5 q_p \sqrt{n q_c},$$

on obtient que

$$\Pr\left[\widetilde{P}_0 \leftarrow_{\$} \mathsf{TP}(\mathcal{T}, n), \mathbf{P} \leftarrow_{\$} (\mathsf{P}(n))^4 \,:\, |\mathsf{BadK}_1| > C\right] \leq \frac{2}{N}.$$

Utiliser (4.16) nous donne

$$\Pr\left[T_{\mathrm{id}} \in \Theta_1\right] \leq \frac{q_c q_p^2}{N^2} + \frac{2 q_c^2 q_p}{N^2} + \frac{5 q_p \sqrt{n q_c}}{N} + \frac{2}{N}.$$

Conditions (B-2) et (B-3). Considérons la condition (B-2). Pour tout $(t, x, y) \in \mathcal{Q}_C$, tout $(u_1, v_1) \in \mathcal{Q}_{P_1}$, et tout $(u_2, v_2) \in \mathcal{Q}_{P_2}$, la probabilité, sur le choix aléatoire de (k_0, k_1), que $k_0 \oplus t = x \oplus u_1$ et $k_1 \oplus t = v_1 \oplus u_2$ vaut $1/N^2$ puisque (k_0, k_1) est tirée uniformément et indépendamment de la transcription des requêtes. Sommer sur les $q_c q_p^2$ choix possibles pour (t, x, y), (u_1, v_1), et (u_2, v_2) donne

$$\Pr\left[T_{\mathrm{id}} \in \Theta_2\right] \leq \frac{q_c q_p^2}{N^2}.$$

De même,

$$\Pr\left[T_{\mathrm{id}} \in \Theta_3\right] \leq \frac{q_c q_p^2}{N^2}.$$

Conditions (B-4) et (B-5). Considérons la condition (B-4). En voyant α_1 comme une variable aléatoire sur le choix uniformément aléatoire de (k_0, k_1), on a

$$\mathbb{E}[\alpha_1] = \sum_{(t,x,y) \in \mathcal{Q}_C} \sum_{u_1 \in U_1} \Pr\left[k_0 = x \oplus u_1 \oplus t\right] \leq \frac{q_c q_p}{N}.$$

Par conséquent, d'après l'inégalité de Markov,

$$\Pr\left[T_{\mathrm{id}} \in \Theta_4\right] = \Pr\left[\alpha_1 \geq \frac{\sqrt{q_c}}{2}\right] \leq \frac{2\mathbb{E}[\alpha_1]}{\sqrt{q_c}} \leq \frac{2q_p\sqrt{q_c}}{N}.$$

De même,

$$\Pr\left[T_{\mathrm{id}} \in \Theta_5\right] \leq \frac{2q_p\sqrt{q_c}}{N}.$$

Condition (B-6). Comme précédemment, voyons $\alpha_{2,3}$ comme une variable aléatoire sur le choix aléatoire de k_0. Ainsi

$$\mathbb{E}[\alpha_{2,3}] = \sum_{(t,x,y)\in\mathcal{Q}_C} \sum_{v_2\in V_2} \sum_{u_3\in U_3} \Pr\left[k_0 = v_2 \oplus u_3 \oplus t\right] \leq \frac{q_c q_p^2}{N}.$$

Donc, d'après l'inégalité de Markov,

$$\Pr\left[T_{\mathrm{id}} \in \Theta_6\right] = \Pr\left[\alpha_{2,3} \geq q_p\sqrt{q_c}\right] \leq \frac{\mathbb{E}[\alpha_{2,3}]}{q_p\sqrt{q_c}} \leq \frac{q_p\sqrt{q_c}}{N}.$$

Conditions (B-7) et (B-8). Considérons la condition (B-7). On considère le distingueur combiné avec $\widetilde{P}_0$ et les permutations internes P_1, P_3, et P_4 comme un algorithme probabiliste $\mathcal{A}$ interagissant avec P_2, et on voit ν_2 comme une variable aléatoire sur le choix aléatoire de P_2 et l'aléa de $\mathcal{A}$. On a

$$\mathbb{E}[\nu_2] = \sum_{\substack{(i,j)\\ 1\leq i\neq j\leq q_c}} \Pr\left[u_{2,i} \oplus v_{2,i} = u_{2,j} \oplus v_{2,j}\right],$$

où les requêtes à P_2 sont rangées dans le même ordre que celui dans lequel elles ont été effectuées par $\mathcal{A}$. Considérons les $i^{\text{ème}}$ et $j^{\text{ème}}$ requêtes, et supposons sans perte de généralité que $i < j$. Si la $j^{\text{ème}}$ requête $u_{2,j}$ est directe, alors $v_{2,j}$ est uniformément aléatoire dans un ensemble de taille $N - j + 1$. De même, s'il s'agit d'une requête inverse $v_{2,j}$, alors $u_{2,j}$ est uniformément aléatoire dans un ensemble de taille $N - j + 1$. Dans tous les cas, la probabilité que $u_{2,i} \oplus v_{2,i} = u_{2,j} \oplus v_{2,j}$ vaut au plus $1/(N - q_p)$. Par conséquent,

$$\mathbb{E}[\nu_2] \leq \frac{q_p(q_p - 1)}{N - q_p} \leq \frac{2q_p^2}{N}.$$

D'après l'inégalité de Markov,

$$\Pr\left[T_{\mathrm{id}} \in \Theta_7\right] = \Pr\left[\nu_2 \geq \sqrt{q_p}\right] \leq \frac{2q_p^{3/2}}{N}.$$

De même,

$$\Pr\left[T_{\mathrm{id}} \in \Theta_8\right] \leq \frac{2q_p^{3/2}}{N}.$$

Le résultat se déduit grâce à l'inégalité de Boole. □

4.4.4 Étude des bonnes transcriptions

Dans cette sous-section, on fixe une bonne transcription $\tau = (\mathcal{Q}_C, \mathcal{Q}_{P_1}, \ldots, \mathcal{Q}_{P_4}, (k_0, k_1))$. D'après (2.4), nous avons à minorer

$$\mathsf{p}(\tau) \stackrel{\text{def}}{=} \Pr\left[\mathbf{P} \leftarrow_{\$} (\mathsf{P}(n))^4 : \mathsf{TEM}^{\mathbf{P}}_{k_0,k_1} \vdash \mathcal{Q}_C \middle| P_1 \vdash \mathcal{Q}_{P_1} \wedge \ldots \wedge P_4 \vdash \mathcal{Q}_{P_4}\right].$$

La preuve comportera deux étapes : tout d'abord, nous minorerons la probabilité que les permutations P_1 et P_4 vérifient des conditions définies dans la définition ci-dessous, et ensuite, en supposant que (P_1, P_4) ne vérifient pas ces conditions, nous minorerons la probabilité, sur les choix aléatoires de P_2 et P_3, que $\mathsf{TEM}^{\mathbf{P}}_{k_0,k_1} \vdash \mathcal{Q}_C$. Pour cette seconde étape, nous pourrons directement réutiliser le lemme 17.

Commençons par donner les conditions qui définissent une bonne paire de permutations (P_1, P_4). Il faut insister sur le fait que ces conditions ne peuvent pas être intégrées dans la définition des mauvaises transcriptions puisqu'elles dépendent de valeurs de P_1 et P_4 qui n'apparaissent pas dans la transcription des requêtes, et qu'elles ne peuvent donc pas être définies uniquement à partir de la transcription τ. Remarquons également que les conditions (C-5) et (C-6) sont factices et on prouvera aisément qu'elles ne peuvent pas être satisfaites. Toutefois elles nous permettront d'utiliser proprement le lemme 17.

Définition 12. *Une paire de permutations (P_1, P_4) telles que $P_1 \vdash \mathcal{Q}_{P_1}$ and $P_4 \vdash \mathcal{Q}_{P_4}$ est dite mauvaise si au moins l'une des conditions suivantes est vérifiée (voir la figure 4.3 pour un diagramme représentant les dix premières conditions) :*

(C-1) il existe $(t, x, y) \in \mathcal{Q}_C$, $u_2 \in U_2$, et $v_3 \in V_3$ tels que

$$\begin{cases} P_1(x \oplus k_0 \oplus t) \oplus k_1 \oplus t = u_2 \\ P_4^{-1}(y \oplus k_0 \oplus t) \oplus k_1 \oplus t = v_3; \end{cases}$$

(C-2) il existe $(t, x, y) \in \mathcal{Q}_C$, $(u_2, v_2) \in \mathcal{Q}_{P_2}$, et $u_3 \in U_3$ tels que

$$\begin{cases} P_1(x \oplus k_0 \oplus t) \oplus k_1 \oplus t = u_2 \\ v_2 \oplus k_0 \oplus t = u_3; \end{cases}$$

(C-3) il existe $(t, x, y) \in \mathcal{Q}_C$, $(u_3, v_3) \in \mathcal{Q}_{P_3}$, et $v_2 \in V_2$ tels que

$$\begin{cases} P_4^{-1}(y \oplus k_0 \oplus t) \oplus k_1 \oplus t = v_3 \\ u_3 \oplus k_0 \oplus t = v_2; \end{cases}$$

(C-4) il existe $(t, x, y), (t', x', y'), (t'', x'', y'') \in \mathcal{Q}_C$, (t, x, y) étant différent de (t', x', y') et de (t'', x'', y'') tels que

$$\begin{cases} P_1(x \oplus k_0 \oplus t) \oplus t = P_1(x' \oplus k_0 \oplus t') \oplus t' \\ P_4^{-1}(y \oplus k_0 \oplus t) \oplus t = P_4^{-1}(y'' \oplus k_0 \oplus t'') \oplus t''; \end{cases}$$

(C-5) il existe $(t, x, y,) \neq (t', x', y') \in \mathcal{Q}_C$ tels que

$$\begin{cases} P_1(x \oplus k_0 \oplus t) \oplus t = P_1(x' \oplus k_0 \oplus t') \oplus t' \\ t = t'; \end{cases}$$

(C-6) il existe $(t,x,y,) \neq (t',x',y') \in \mathcal{Q}_C$ *tels que*

$$\begin{cases} P_4^{-1}(y \oplus k_0 \oplus t) \oplus t = P_4^{-1}(y' \oplus k_0 \oplus t') \oplus t' \\ t = t'; \end{cases}$$

(C-7) il existe $(t,x,y) \neq (t',x',y') \in \mathcal{Q}_C$ *et* $u_2 \in U_2$ *tels que*

$$\begin{cases} P_1(x \oplus k_0 \oplus t) \oplus k_1 \oplus t = u_2 \\ P_4^{-1}(y \oplus k_0 \oplus t) \oplus t = P_4^{-1}(y' \oplus k_0 \oplus t') \oplus t'; \end{cases}$$

(C-8) il existe $(t,x,y) \neq (t',x',y') \in \mathcal{Q}_C$ *et* $v_3 \in V_3$ *tels que*

$$\begin{cases} P_4^{-1}(y \oplus k_0 \oplus t) \oplus k_1 \oplus t = v_3 \\ P_1(x \oplus k_0 \oplus t) \oplus t = P_1(x' \oplus k_0 \oplus t') \oplus t'; \end{cases}$$

(C-9) il existe $(t,x,y) \neq (t',x',y') \in \mathcal{Q}_C$ *et* $(u_2,v_2),(u_2',v_2') \in \mathcal{Q}_{P_2}$ *tels que*

$$\begin{cases} P_1(x \oplus k_0 \oplus t) \oplus k_1 \oplus t = u_2 \\ P_1(x' \oplus k_0 \oplus t') \oplus k_1 \oplus t' = u_2' \\ v_2 \oplus t = v_2' \oplus t'; \end{cases}$$

(C-10) il existe $(t,x,y) \neq (t',x',y') \in \mathcal{Q}_C$ *et* $(u_3,v_3),(u_3',v_3') \in \mathcal{Q}_{P_3}$ *tels que*

$$\begin{cases} P_4^{-1}(y \oplus k_0 \oplus t) \oplus k_1 \oplus t = v_3 \\ P_4^{-1}(y' \oplus k_0 \oplus t') \oplus k_1 \oplus t' = v_3' \\ u_3 \oplus t = u_3' \oplus t'; \end{cases}$$

(C-11) $\alpha_2 \geq \sqrt{q_c}$ *;*
(C-12) $\alpha_3 \geq \sqrt{q_c}$ *;*
(C-13) $\beta_2 \geq \sqrt{q_c}$ *;*
(C-14) $\beta_3 \geq \sqrt{q_c}$ *;*
où

$$\begin{aligned}
\alpha_2 &\stackrel{\text{def}}{=} |\{(t,x,y) \in \mathcal{Q}_C \,:\, P_1(x \oplus k_0 \oplus t) \oplus k_1 \oplus t \in U_2\}|, \\
\alpha_3 &\stackrel{\text{def}}{=} |\{(t,x,y) \in \mathcal{Q}_C \,:\, P_4^{-1}(y \oplus k_0 \oplus t) \oplus k_1 \oplus t \in V_3\}|, \\
\beta_2 &\stackrel{\text{def}}{=} |\{(t,x,y) \in \mathcal{Q}_C \,:\, \exists (t',x',y') \neq (t,x,y), \\
&\qquad P_1(x \oplus k_0 \oplus t) \oplus t = P_1(x' \oplus k_0 \oplus t') \oplus t'\}|, \\
\beta_3 &\stackrel{\text{def}}{=} |\{(t,x,y) \in \mathcal{Q}_C \,:\, \exists (t',x',y') \neq (t,x,y), \\
&\qquad P_4^{-1}(y \oplus k_0 \oplus t) \oplus t = P_4^{-1}(y' \oplus k_0 \oplus t') \oplus t'\}|.
\end{aligned}$$

Dans le cas contraire, (P_1,P_4) *est une bonne paire de permutations. On note* Π_{good}*, respectivement* Π_{bad} *les ensemble des bonnes, respectivement des mauvaises paires de permutations* (P_1,P_4) *telles que* $P_1 \vdash \mathcal{Q}_{P_1}$ *et* $P_4 \vdash \mathcal{Q}_{P_4}$*.*

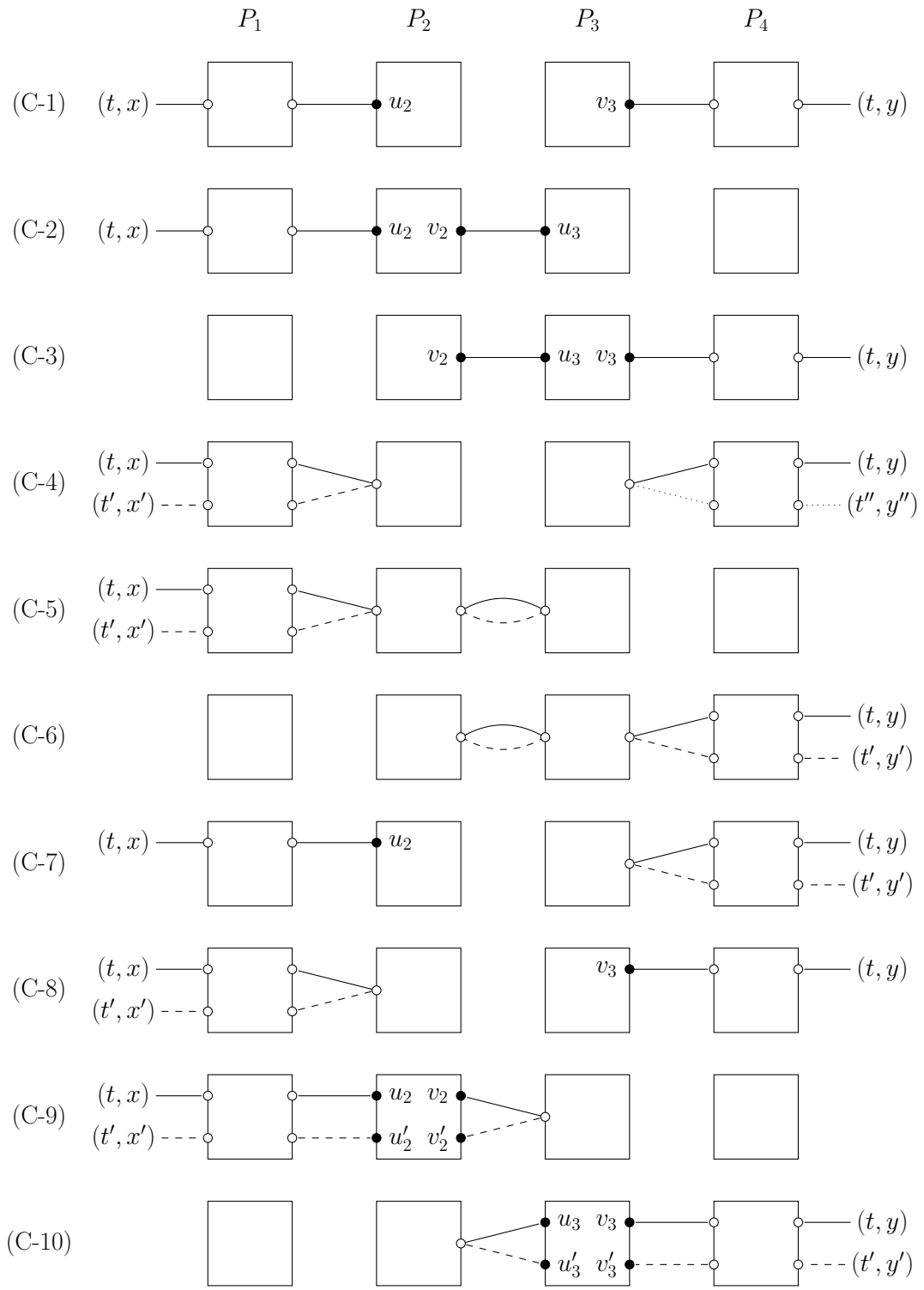

FIGURE 4.3 – Les dix conditions de « collision » qui caractérisent une mauvaise paire de permutations (P_1, P_4). Les points noirs correspondent aux paires $(u_2, v_2) \in \mathcal{Q}_{P_2}$ ou $(u_3, v_3) \in \mathcal{Q}_{P_3}$. Notons que, pour (C-4), on peut avoir $(t', x') = (t'', x'')$, et pour (C-9) (respectivement (C-10)) on peut avoir $x \oplus t = x' \oplus t'$ (respectivement $y \oplus t = y' \oplus t'$).

Dans le reste de la preuve, on note Π l'ensemble des paires de permutations (P_1, P_4) telles que $P_1 \vdash \mathcal{Q}_{P_1}$ et $P_4 \vdash \mathcal{Q}_{P_4}$. La première étape de l'étude des bonnes transcriptions sera la majoration de la probabilité que la paire (P_1, P_4) soit mauvaise.

Lemme 30. *Pour toute paire d'entiers naturels (q_c, q_p) tels que $q_p + q_c + 1 \leq N/2$, on a*

$$\Pr[(P_1, P_4) \in \Pi_{\text{bad}}] \leq \frac{4q_c^3 + 16q_c^2 q_p + 4q_c q_p^2}{N^2} + \frac{10q_c^{3/2} + 4q_c\sqrt{q_p} + 10\sqrt{q_c}q_p}{N}$$

où la probabilité est prise sur le choix uniformément aléatoire de (P_1, P_4) dans Π.

Démonstration. Nous allons majorer successivement les probabilités d'obtenir chacune des quatorze conditions. Soit Π_i l'ensemble des paires de permutations $(P_1, P_4) \in \Pi$ qui vérifient la condition (C-i).

Condition (C-1). Fixons $(t, x, y) \in \mathcal{Q}_C$, $u_2 \in U_2$, et $v_3 \in V_3$. D'une part, notons que si $x \oplus k_0 \oplus t = u_1$ pour une requête $(u_1, v_1) \in \mathcal{Q}_{P_1}$, alors $v_1 \oplus k_1 \oplus t$ ne peut pas être égal à u_2 car sinon τ vérifierait la condition (B-2). De même, si $y \oplus k_0 \oplus t = v_4$ pour une requête $(u_4, v_4) \in \mathcal{Q}_{P_4}$, alors $u_4 \oplus k_1 \oplus t$ ne peut pas être égal à v_3 car sinon τ vérifierait (B-3). D'autre part, si $x \oplus k_0 \oplus t \notin U_1$ et $y \oplus k_0 \oplus t \notin V_4$, alors la probabilité quand $(P_1, P_4) \leftarrow_\$ \Pi$ que

$$\begin{cases} P_1(x \oplus k_0 \oplus t) = u_2 \oplus k_1 \oplus t \\ P_4^{-1}(y \oplus k_0 \oplus t) = v_3 \oplus k_1 \oplus t \end{cases}$$

est au plus $1/(N - q_p)^2 \leq 4/N^2$. En effet, si $u_2 \oplus k_1 \oplus t \in V_1$ or $v_3 \oplus k_1 \oplus t \in U_4$, alors cette probabilité est nulle, alors que sinon elle vaut exactement $1/(N - q_p)^2$. En sommant sur les au plus $q_c q_p^2$ choix possibles pour (t, x, y), u_2, et v_3, on obtient

$$\Pr[(P_1, P_4) \in \Pi_1] \leq \frac{4q_c q_p^2}{N^2}.$$

Conditions (C-2) et (C-3). Considérons (C-2), le raisonnement pour (C-3) étant similaire. Fixons $(t, x, y) \in \mathcal{Q}_C$, $(u_2, v_2) \in \mathcal{Q}_{P_2}$, et $u_3 \in U_3$. Notons tout d'abord que, pour que (C-2) soit satisfaite, on doit avoir $v_2 \oplus k_0 \oplus t = u_3$, et il y a par définition au plus $\alpha_{2,3}$ triplets $((t, x, y), v_2, u_3)$ qui vérifient cette égalité. Si $x \oplus k_0 \oplus t = u_1$ pour une requête $(u_1, v_1) \in \mathcal{Q}_{P_1}$, alors $v_1 \oplus k_1 \oplus t$ ne peut pas être égale à u_2 car sinon τ vérifierait (B-2). De plus, si $x \oplus k_0 \oplus t \notin U_1$, alors la probabilité que $P_1(x \oplus k_0 \oplus t) = u_2 \oplus k_1 \oplus t$ vaut au plus $1/(N - q_p) \leq 2/N$. En effet, elle est nulle si $u_2 \oplus k_1 \oplus t \in V_1$, et vaut $1/(N - q_p)$ sinon. En sommant sur les au plus $\alpha_{2,3}$ possibilités pour (t, x, y), (u_2, v_2), et u_3, en remarquant que $\alpha_{2,3} \leq q_p\sqrt{q_c}$ car sinon τ vérifierait (B-6), on obtient

$$\Pr[(P_1, P_4) \in \Pi_2] \leq \frac{2q_p\sqrt{q_c}}{N}.$$

De même,

$$\Pr[(P_1, P_4) \in \Pi_3] \leq \frac{2q_p\sqrt{q_c}}{N}.$$

Condition (C-4). Fixons $(t,x,y),(t',x',y'),(t'',x'',y'') \in \mathcal{Q}_C$ tels que (t,x,y) est différent de (t',x',y') et de (t'',x'',y''). Tout d'abord, remarquons que si $x \oplus k_0 \oplus t = x' \oplus k_0 \oplus t'$ ou $y \oplus k_0 \oplus t = y'' \oplus k_0 \oplus t''$, alors (C-4) ne peut pas être satisfaite. Par conséquent, supposons qu'aucune de ces deux égalités ne soit vérifiée. On considère trois cas. Supposons d'abord que $x \oplus k_0 \oplus t = u_1$ pour une requête $(u_1,v_1) \in \mathcal{Q}_{P_1}$. Remarquons qu'il y a au plus α_1 possibilités pour (t,x,y), or $\alpha_1 \leq \sqrt{q_c}/2$ car sinon τ vérifierait (B-4). De plus $y \oplus k_0 \oplus t \notin V_4$ car sinon τ vérifierait (B-1). Ainsi, la probabilité que

$$P_4^{-1}(y \oplus k_0 \oplus t) \oplus t = P_4^{-1}(y'' \oplus k_0 \oplus t'') \oplus t''$$

vaut au plus $1/(N-q_p-1) \leq 2/N$. En effet, si $y'' \oplus k_0 \oplus t'' \in V_4$, alors cette probabilité vaut soit zéro si $P_4^{-1}(y'' \oplus k_0 \oplus t'') \oplus t \oplus t'' \in U_4$, soit exactement $1/(N-q_p)$ dans le cas contraire, alors que si $y'' \oplus k_0 \oplus t'' \notin V_4$, alors cette probabilité est inférieure à $1/(N-q_p-1)$. En sommant sur les au plus $\sqrt{q_c}/2 \times q_c$ choix possibles pour (t,x,y) et (t'',x'',y''), la probabilité dans ce premier cas vaut au plus $q_c^{3/2}/N$. Le deuxième cas dans lequel $y \oplus k_0 \oplus t \in V_4$ se traite de manière similaire. Enfin, considérons le cas pour lequel $x \oplus k_0 \oplus t \notin U_1$ et $y \oplus k_0 \oplus t \notin V_4$. Ici, la probabilité que

$$\begin{cases} P_1(x \oplus k_0 \oplus t) \oplus t = P_1(x' \oplus k_0 \oplus t') \oplus t' \\ P_4^{-1}(y \oplus k_0 \oplus t) \oplus t = P_4^{-1}(y'' \oplus k_0 \oplus t'') \oplus t''; \end{cases}$$

vaut au plus $1/(N-q_p-1)^2 \leq 4/N^2$. En sommant sur les au plus q_c^3 choix possibles pour (t,x,y), (t',x',y'), et (t'',x'',y''), la probabilité de ce dernier cas vaut au plus $4q_c^3/N^2$. Nous obtenons donc en tout que

$$\Pr[(P_1,P_4) \in \Pi_4] \leq \frac{4q_c^3}{N^2} + \frac{2q_c^{3/2}}{N}.$$

Conditions (C-5) et (C-6). Ces conditions ne peuvent pas être vérifiées. En effet, supposons qu'il existe $(t,x,y) \neq (t',x',y') \in \mathcal{Q}_C$ satisfaisant (C-5). Puisque $t = t'$, alors $x \neq x'$ d'après l'hypothèse que le distingueur n'effectue jamais de requête inutile. Ceci entraîne évidemment que $P_1(x \oplus k_0 \oplus t) \oplus t \neq P_1(x' \oplus k_0 \oplus t') \oplus t'$, ce qui est contradictoire. Le raisonnement est similaire pour (C-6). Ainsi,

$$\Pr[(P_1,P_4) \in \Pi_5] = \Pr[(P_1,P_4) \in \Pi_6] = 0.$$

Conditions (C-7) et (C-8). Considérons la condition (C-7). Fixons des requêtes $(t,x,y) \neq (t',x',y') \in \mathcal{Q}_C$ et $u_2 \in U_2$. Nous allons considérer deux cas : le cas dans lequel $y \oplus k_0 \oplus t \in V_4$, et celui dans lequel $y \oplus k_0 \oplus t \notin V_4$. Dans les deux cas, remarquons que si $x \oplus k_0 \oplus t = u_1$ pour une requête $(u_1,v_1) \in \mathcal{Q}_{P_1}$, alors $v_1 \oplus k_1 \oplus t$ ne peut pas être égal à u_2 car sinon τ vérifierait (B-2). Ainsi, on peut supposer que $x \oplus k_0 \oplus t \notin U_1$. Il en résulte que la probabilité que

$$P_1(x \oplus k_0 \oplus t) \oplus k_1 \oplus t = u_2$$

vaut au plus $1/(N-q_p) \leq 2/N$. En effet, elle est nulle si $u_2 \oplus k_1 \oplus t \in V_1$, et vaut $1/(N-q_p)$ sinon. En sommant sur les au plus α_4 requêtes $(t,x,y) \in \mathcal{Q}_C$ telles que

$y \oplus k_0 \oplus t \in V_4$, et en remarquant que $\alpha_4 \leq \sqrt{q_c}/2$ car sinon τ vérifierait (B-5), et les q_p choix possibles pour u_2, on constate que le premier cas se produit avec une probabilité inférieure à $q_p\sqrt{q_c}/N$. Supposons maintenant que $y \oplus k_0 \oplus t \notin V_4$. Alors la probabilité que

$$P_4^{-1}(y \oplus k_0 \oplus t) \oplus t = P_4^{-1}(y' \oplus k_0 \oplus t') \oplus t'$$

est inférieure à $1/(N - q_p - 1) \leq 2/N$. En effet, si $y \oplus k_0 \oplus t = y' \oplus k_0 \oplus t'$, alors il est clair que cette égalité ne peut pas être satisfaite, alors que si $y \oplus k_0 \oplus t \neq y' \oplus k_0 \oplus t'$, l'égalité est vraie avec une probabilité inférieure à $1/(N - q_p - 1)$. En sommant sur les au plus $q_c^2 q_p$ choix possibles pour (t, x, y), (t', x', y'), et u_2, on montre que la probabilité de ce second cas est inférieure à $4q_c^2 q_p/N^2$. Dans l'ensemble,

$$\Pr\left[(P_1, P_4) \in \Pi_7\right] \leq \frac{q_p\sqrt{q_c}}{N} + \frac{4q_c^2 q_p}{N^2}.$$

De même, on a

$$\Pr\left[(P_1, P_4) \in \Pi_8\right] \leq \frac{q_p\sqrt{q_c}}{N} + \frac{4q_c^2 q_p}{N^2}.$$

Conditions (C-9) et (C-10). Considérons la condition (C-9). Commençons par remarquer que, si cette condition est satisfaite, on a $x \oplus k_0 \oplus t \notin U_1$, $x' \oplus k_0 \oplus t' \notin U_1$, $u_2 \oplus k_1 \oplus t \notin V_1$ et $u_2' \oplus k_1 \oplus t' \notin V_1$, car sinon (B-2) serait vérifiée. De plus, si $(u_2, v_2) = (u_2', v_2')$, alors $t = t'$, d'où $x = x'$, ce qui est impossible. Ainsi on doit avoir $(u_2, v_2) \neq (u_2', v_2')$. La condition peut alors être décomposée en deux sous-conditions :

9.1 il existe $(t, x, y) \neq (t', x', y') \in \mathcal{Q}_C$ et $(u_2, v_2) \neq (u_2', v_2') \in \mathcal{Q}_{P_2}$ tels que $x \oplus t = x' \oplus t'$, $P_1(x \oplus k_0 \oplus t) = u_2 \oplus k_1 \oplus t$ et $P_1(x' \oplus k_0 \oplus t') = u_2' \oplus k_1 \oplus t'$ et $v_2 \oplus t = v_2' \oplus t'$;

9.2 il existe $(t, x, y) \neq (t', x', y') \in \mathcal{Q}_C$ et $(u_2, v_2) \neq (u_2', v_2') \in \mathcal{Q}_{P_2}$ tels que $x \oplus t \neq x' \oplus t'$, $P_1(x \oplus k_0 \oplus t) = u_2 \oplus k_1 \oplus t$ et $P_1(x' \oplus k_0 \oplus t') = u_2' \oplus k_1 \oplus t'$ et $v_2 \oplus t = v_2' \oplus t'$.

Dans le premier cas, on a

$$u_2 \oplus k_1 \oplus t = P_1(x \oplus k_0 \oplus t) = P_1(x' \oplus k_0 \oplus t') = u_2' \oplus k_1 \oplus t',$$

d'où $u_2 \oplus u_2' = t \oplus t' = v_2 \oplus v_2'$. Ainsi la première condition entraîne la suivante : il existe $(t, x, y) \in \mathcal{Q}_C$ et $(u_2, v_2) \neq (u_2', v_2') \in \mathcal{Q}_{P_2}$ tels que $P_1(x \oplus k_0 \oplus t) = u_2 \oplus k_1 \oplus t$ et $u_2 \oplus u_2' = v_2 \oplus v_2'$, où $x \oplus k_0 \oplus t \notin U_1$ et $u_2 \oplus k_1 \oplus t \notin V_1$. Puisque $\nu_2 < \sqrt{q_p}$, le nombre de $u_2 \in U_2$ convenables est inférieur à $\sqrt{q_p}$, et la probabilité que cette première condition soit vérifiée est inférieure à $\frac{q_c\sqrt{q_p}}{N - q_p} \leq \frac{2q_c\sqrt{q_p}}{N}$. En ce qui concerne la seconde condition, fixons des requêtes quelconques $(t, x, y) \neq (t', x', y') \in \mathcal{Q}_C$ telles que $x \oplus t \neq x' \oplus t'$, $x \oplus k_0 \oplus t \notin U_1$, $x' \oplus k_0 \oplus t' \notin U_1$ et $(u_2, v_2) \in \mathcal{Q}_{P_2}$. Si $v_2 \oplus t \oplus t' \notin V_2$, la condition ne peut pas être satisfaite. Dans le cas contraire, soit $(u_2', v_2') \in \mathcal{Q}_{P_2}$ l'unique requête telle que $v_2 \oplus t = v_2' \oplus t'$. Alors la probabilité que $P_1(x \oplus k_0 \oplus t) = u_2 \oplus k_1 \oplus t$ et $P_1(x' \oplus k_0 \oplus t') = u_2' \oplus k_1 \oplus t'$ est inférieure à $\frac{1}{(N - q_p)(N - q_p - 1)}$. Ainsi, en sommant sur tous les uplets de requêtes possibles, et en tenant compte de la condition 9.1, on a

$$\Pr\left[(P_1, P_4) \in \Pi_9\right] \leq \frac{2q_c\sqrt{q_p}}{N} + \frac{4q_c^2 q_p}{N^2}.$$

De même,

$$\Pr\left[(P_1, P_4) \in \Pi_{10}\right] \leq \frac{2q_c\sqrt{q_p}}{N} + \frac{4q_c^2 q_p}{N^2}.$$

Conditions (C-11) et (C-12). Voyons α_2 (respectivement α_3) comme une variable aléatoire sur le choix de P_1 (respectivement P_4). Remarquons que

$$\begin{aligned}\alpha_2 &= |\{(t,x,y) \in \mathcal{Q}_C \,:\, P_1(x \oplus k_0 \oplus t) \oplus k_1 \oplus t \in U_2\}| \\ &= |\{(t,x,y) \in \mathcal{Q}_C \,:\, x \oplus k_0 \oplus t \notin U_1,\, P_1(x \oplus k_0 \oplus t) \oplus k_1 \oplus t \in U_2\}|,\end{aligned}$$

car, si $x \oplus k_0 \oplus t \in U_1$ et $P_1(x \oplus k_0 \oplus t) \oplus k_1 \oplus t \in U_2$, alors (B-2) est vérifiée. On note $\mathcal{Q}_{C,1}$ le sous-ensemble des requêtes $(t,x,y) \in \mathcal{Q}_C$ telles que $x \oplus k_0 \oplus t \notin U_1$. Alors

$$\begin{aligned}\mathbb{E}[\alpha_2] &= \sum_{(t,x,y)\in\mathcal{Q}_{C,1}} \sum_{u_2 \in U_2} \Pr\left[P_1(x \oplus k_0 \oplus t) = u_2 \oplus k_1 \oplus t\right] \\ &\leq \sum_{(t,x,y)\in\mathcal{Q}_{C,1}} \sum_{u_2 \in U_2} \frac{1}{N - q_p} \\ &\leq \frac{2q_c q_p}{N}.\end{aligned}$$

D'après l'inégalité de Markov, on obtient

$$\Pr\left[(P_1, P_4) \in \Pi_{11}\right] \leq \frac{2q_p\sqrt{q_c}}{N}.$$

De même,

$$\Pr\left[(P_1, P_4) \in \Pi_{12}\right] \leq \frac{2q_p\sqrt{q_c}}{N}.$$

Conditions (C-13) et (C-14). Considérons la condition (C-13). On remarque que

$$\begin{aligned}\beta_2 = |\{&(t,x,y) \in \mathcal{Q}_C \,:\, \exists (t',x',y') \neq (t,x,y), \\ &P_1(x \oplus k_0 \oplus t) \oplus t = P_1(x' \oplus k_0 \oplus t') \oplus t'\}| \\ \leq \alpha_1 + |\{&(t,x,y) \in \mathcal{Q}_C \,:\, x \oplus k_0 \oplus t \notin U_1 \text{ et } \exists (t',x',y') \neq (t,x,y), \\ &P_1(x \oplus k_0 \oplus t) \oplus t = P_1(x' \oplus k_0 \oplus t') \oplus t'\}|.\end{aligned}$$

Notons β_2' le dernier terme de cette somme. Ainsi

$$\begin{aligned}\mathbb{E}[\beta_2'] &= \sum_{(t,x,y)\in\mathcal{Q}_{C,1}} \sum_{(t',x',y')\neq(t,x,y)} \Pr\left[P_1(x \oplus k_0 \oplus t) \oplus t = P_1(x' \oplus k_0 \oplus t') \oplus t'\right] \\ &\leq \frac{q_c^2}{N - q_p - 1} \leq \frac{2q_c^2}{N}.\end{aligned}$$

Cette inégalité est vraie car, si $x \oplus t = x' \oplus t'$, alors $t \neq t'$ puisque le distingueur n'effectue jamais de requête inutile, donc $P_1(x \oplus k_0 \oplus t) \oplus t = P_1(x' \oplus k_0 \oplus t') \oplus t'$ ne peut pas être satisfaite. Dans le cas contraire,

$$\Pr\left[P_1(x \oplus k_0 \oplus t) \oplus t = P_1(x' \oplus k_0 \oplus t') \oplus t'\right] \leq \frac{1}{N - q_p - 1}.$$

Or, puisque (B-4) n'est pas satisfaite, $\alpha_1 < \sqrt{q_c}/2$. Ainsi $\beta_2 \geq \sqrt{q_c}$ entraîne $\beta'_2 \geq \sqrt{q_c}/2$. D'où, d'après l'inégalité de Markov,

$$\Pr\left[(P_1, P_4) \in \Pi_{13}\right] \leq \Pr\left[\beta'_2 \geq \sqrt{q_c}/2\right] \leq \frac{2\mathbb{E}[\beta'_2]}{\sqrt{q_c}} \leq \frac{4q_c^{3/2}}{N}.$$

De même,

$$\Pr\left[(P_1, P_4) \in \Pi_{14}\right] \leq \frac{4q_c^{3/2}}{N}.$$

L'inégalité de Boole nous permet d'en déduire le résultat. □

Nous pouvons à présent entamer la seconde partie du raisonnement.

Définition 13. *Fixons une paire quelconque de permutations (P_1, P_4) telles que $P_1 \vdash \mathcal{Q}_{P_1}$ et $P_4 \vdash \mathcal{Q}_{P_4}$. Définissons une nouvelle transcription des requêtes $\mathcal{Q}'_C$ qui dépend du choix de (P_1, P_4) de la façon suivante*

$$\mathcal{Q}'_C = \{(t, P_1(x \oplus k_0 \oplus t), P_4^{-1}(y \oplus k_0 \oplus t)) : (t, x, y) \in \mathcal{Q}_C\}.$$

On définit également

$$\tilde{\mathsf{p}}(\tau, P_1, P_4) = \Pr\left[P_2, P_3 \leftarrow_{\$} \mathsf{P}(n) : \mathsf{TEM}_{k_1,k_0}^{P_2,P_3} \vdash \mathcal{Q}'_C \,\middle|\, (P_2 \vdash \mathcal{Q}_{P_2}) \wedge (P_3 \vdash \mathcal{Q}_{P_3})\right].$$

Lemme 31. *On a*

$$\frac{\Pr\left[T_{\mathrm{re}} = \tau\right]}{\Pr\left[T_{\mathrm{id}} = \tau\right]} \geq \sum_{(P_1,P_4) \in \Pi_{\mathrm{good}}} \frac{\tilde{\mathsf{p}}(\tau, P_1, P_4)}{\left((N-q_p)!\right)^2 \prod_{i=1}^{m} 1/(N)_{q_i}}.$$

Démonstration. Il est clair que, une fois P_1 et P_4 fixées, $\mathsf{TEM}_{k_0,k_1}^{P_1,P_2,P_3,P_4} \vdash \mathcal{Q}_C$ équivaut à $\mathsf{TEM}_{k_1,k_0}^{P_2,P_3} \vdash \mathcal{Q}'_C$. Ainsi,

$$\begin{aligned}
\mathsf{p}(\tau) &= \sum_{(\bar{P}_1,\bar{P}_4) \in \Pi} \Pr\left[(P_1, P_4) \leftarrow_{\$} \Pi : (P_1 = \bar{P}_1) \wedge (P_4 = \bar{P}_4)\right] \tilde{\mathsf{p}}(\tau, \bar{P}_1, \bar{P}_4) \\
&\geq \sum_{(\bar{P}_1,\bar{P}_4) \in \Pi_{\mathrm{good}}} \frac{\tilde{\mathsf{p}}(\tau, \bar{P}_1, \bar{P}_4)}{((N-q_p)!)^2}.
\end{aligned}$$

On obtient le résultat grâce à l'équation (2.4). □

On peut maintenant appliquer directement le lemme 17.

Lemme 32. *Soient q_c et q_p deux entiers naturels tels que $q_p + 3q_c \leq N/2$. Fixons une paire quelconque de permutations $(P_1, P_4) \in \Pi_{\mathrm{good}}$. Alors*

$$\frac{\tilde{\mathsf{p}}(\tau, P_1, P_4)}{\prod_{i=1}^{m} 1/(N)_{q_i}} \geq 1 - \left(\frac{4q_c(q_p + 2q_c)^2}{N^2} + \frac{14q_c^{3/2} + 4\sqrt{q_c}q_p}{N}\right).$$

Démonstration. On peut vérifier que la transcription des requêtes $\tau' = (\mathcal{Q}'_C, \mathcal{Q}_{P_2}, \mathcal{Q}_{P_3})$ satisfait exactement les conditions définissant une bonne transcription (voir section 3.3.2). De plus, le rapport $\tilde{\mathsf{p}}(\tau, P_1, P_4)/\prod_{i=1}^{m} 1/(N)_{q_i}$ correspond exactement au rapport entre les probabilités d'obtenir τ' dans le monde réel et dans le monde idéal une fois qu'une bonne paire de permutations (P_1, P_4) a été fixée. Ainsi, on peut appliquer le lemme 17 qui donne directement le résultat. □

Remarque 5. *Il convient de rappeler que la preuve du lemme 17 ne repose pas sur les propriétés des fonctions de hachage h_1 et h_2 qui apparaissent dans la définition des bonnes transcriptions dans la section 3.3.2*

Nous pouvons maintenant prouver le lemme principal de cette sous-section.

Lemme 33. *Soient q_c et q_p deux entiers naturels tels que $q_p + 3q_c + 1 \leq N/2$. On a*

$$\frac{\Pr\left[T_{\text{re}} = \tau\right]}{\Pr\left[T_{\text{id}} = \tau\right]} \geq 1 - \frac{20q_c^3 + 32q_c^2 q_p + 8q_c q_p^2}{N^2} - \frac{24q_c^{3/2} + 4q_c\sqrt{q_p} + 14\sqrt{q_c}q_p}{N}.$$

Démonstration. D'après les lemmes 31 et 32, on a

$$\begin{aligned}
\frac{\Pr\left[T_{\text{re}} = \tau\right]}{\Pr\left[T_{\text{id}} = \tau\right]} &\geq \sum_{(P_1,P_4)\in\Pi_{\text{good}}} \frac{\tilde{\mathsf{p}}(\tau, P_1, P_4)}{\left((N-q_p)!\right)^2 \prod_{i=1}^{m} 1/(N)_{q_i}} \\
&\geq \left(1 - \frac{4q_c(q_p+2q_c)^2}{N^2} - \frac{14q_c^{3/2} + 4\sqrt{q_c}q_p}{N}\right) \sum_{\Pi_{\text{good}}} \frac{1}{\left((N-q_p)!\right)^2} \\
&= \left(1 - \frac{4q_c(q_p+2q_c)^2}{N^2} - \frac{14q_c^{3/2} + 4\sqrt{q_c}q_p}{N}\right) \frac{|\Pi_{\text{good}}|}{\left((N-q_p)!\right)^2} \\
&= \left(1 - \frac{4q_c(q_p+2q_c)^2}{N^2} - \frac{14q_c^{3/2} + 4\sqrt{q_c}q_p}{N}\right) \Pr\left[(P_1, P_4) \in \Pi_{\text{good}}\right],
\end{aligned}$$

où la dernière probabilité est prise sur le choix aléatoire de (P_1, P_4) dans Π, l'ensemble des paires de permutations telles que $P_1 \vdash \mathcal{Q}_{P_1}$ et $P_4 \vdash \mathcal{Q}_{P_4}$. D'après le lemme 30, on a

$$\begin{aligned}
\frac{\Pr\left[T_{\text{re}} = \tau\right]}{\Pr\left[T_{\text{id}} = \tau\right]} &\geq \left(1 - \frac{4q_c^3 + 16q_c^2 q_p + 4q_c q_p^2}{N^2} - \frac{10q_c^{3/2} + 4q_c\sqrt{q_p} + 10\sqrt{q_c}q_p}{N}\right) \\
&\qquad \times \left(1 - \frac{4q_c(q_p+2q_c)^2}{N^2} - \frac{14q_c^{3/2} + 4\sqrt{q_c}q_p}{N}\right) \\
&\geq 1 - \frac{20q_c^3 + 32q_c^2 q_p + 8q_c q_p^2}{N^2} - \frac{24q_c^{3/2} + 4q_c\sqrt{q_p} + 14\sqrt{q_c}q_p}{N}.
\end{aligned}$$

□

Conclusion

Nous pouvons conclure la preuve du théorème 10. En combinant les lemmes 1, 29, et 33, on a

$$\begin{aligned}\mathbf{Adv}^{\mathrm{cca}}_{\mathsf{TEM}[n,4,\mathbf{f}]}(q_c,q_p) &\leq \frac{2q_c^2q_p+3q_cq_p^2}{N^2}+\frac{(5+5\sqrt{n})\sqrt{q_c}q_p+4q_p^{3/2}+2}{N}\\ &\quad+\frac{20q_c^3+32q_c^2q_p+8q_cq_p^2}{N^2}+\frac{24q_c^{3/2}+4q_c\sqrt{q_p}+14\sqrt{q_c}q_p}{N}\\ &\leq \frac{20q_c^3+34q_c^2q_p+11q_cq_p^2}{N^2}\\ &\quad+\frac{24q_c^{3/2}+4q_c\sqrt{q_p}+(19+5\sqrt{n})\sqrt{q_c}q_p+4q_p^{3/2}+2}{N}.\end{aligned}$$

Puisque le résultat est trivialement vrai lorsque $q_c^3 > N^2$, $q_c^2q_p > N^2$, ou $q_cq_p^2 > N^2$, on peut supposer que $q_c^3 \leq N^2$, $q_c^2q_p \leq N^2$, et $q_cq_p^2 \leq N^2$, de telle sorte que

$$\frac{q_c^3}{N^2} \leq \frac{q_c^{3/2}}{N}, \quad \frac{q_c^2q_p}{N^2} \leq \frac{q_c\sqrt{q_p}}{N}, \quad \text{and} \quad \frac{q_cq_p^2}{N^2} \leq \frac{\sqrt{q_c}q_p}{N}.$$

Ainsi

$$\mathbf{Adv}^{\mathrm{cca}}_{\mathsf{TEM}[n,4,\mathbf{f}]}(q_c,q_p) \leq \frac{44q_c^{3/2}+38q_c\sqrt{q_p}+(30+5\sqrt{n})q_p\sqrt{q_c}+4q_p^{3/2}+2}{N},$$

ce qui achève la preuve du théorème 10.

Conclusion

Dans cette thèse, nous avons présenté une généralisation du schéma d'Even-Mansour qui admet une entrée supplémentaire, appelée *tweak*, dont le but est d'apporter une variabilité à l'algorithme de chiffrement. En particulier, nous avons étudié la sécurité de notre construction dans deux cas :

— lorsque l'algorithme de dérivation de clés est hautement non-linéaire,

— lorsque l'algorithme de dérivation de clés est linéaire.

Pour conclure ce manuscrit, nous allons rapidement comparer le niveau de sécurité offert par notre schéma d'Even-Mansour paramétrable à celui garanti par le schéma d'Even-Mansour classique (non-paramétrable) et donner quelques pistes pour de futurs travaux.

Commençons par le cas classique. Le schéma d'Even-Mansour (non-paramétrable) est prouvé sûr jusqu'à environ $2^{rn/(r+1)}$ requêtes de l'adversaire, où n est la taille du bloc et r le nombre de tours, lorsque les clés et les permutations de tours sont indépendantes et uniformément aléatoires [CS14]. Dans le cas particulier où $r = 2$, le niveau de sécurité prouvée reste similaire lorsque les clés de tours sont dérivées d'une seule clé de n bits et les permutations de tours sont identiques [CLL+14].

Comparons ces résultats à ceux que nous avons obtenus au fil de nos travaux. Notre schéma, équipé d'un algorithme de dérivation de clés hautement non-linéaire, lorsque les clés et les permutations de tours sont indépendantes et uniformément aléatoires, offre un niveau de sécurité similaire au cas classique pour un nombre de tours égal à 1 ou 2. Dans le cas général d'un nombre r pair de tours, nous avons démontré que notre construction reste sûre tant que le nombre de requêtes de l'adversaire reste petit devant $2^{rn/(r+2)}$, ce qui correspond au niveau de sécurité du schéma d'Even-Mansour classique à $r/2$ tours.

Dans le cas de l'utilisation d'un algorithme de dérivation de clés linéaire, nous donnons deux attaques très efficaces contre notre construction lorsqu'elle utilisée avec 1 ou 2 tours. Nous présentons ensuite une construction à 3 tours offrant un niveau de sécurité à la borne des anniversaires, puis une construction à 4 tours qui est sûre au-delà de la borne des anniversaires.

L'étude que nous avons menée a également soulevé plusieurs problèmes ouverts intéressants :

1. dans le cas non-linéaire : notre construction offre-t-elle des garanties de sécurité similaires à celles du schéma d'Even-Mansour classique, quel que soit le nombre de tours ?
2. dans le cas linéaire : notre schéma à 4 tours peut-il être généralisé à un nombre arbitraire de tours tout en offrant des garanties de sécurité qui s'améliorent

avec le nombre de tours ?

3. existe-t-il des variantes minimalisées (dans le cas où certaines clés et/ou certaines permutations sont identiques) de nos constructions qui offrent une sécurité comparable à celle de la construction initiale ?

Nous conjecturons que la réponse à ces trois questions est positive. En outre, la technique des coefficients H nous paraît être la méthode la plus adaptée pour le démontrer.

Index

Bibliographie

[ABL+13] Elena Andreeva, Andrey Bogdanov, Atul Luykx, Bart Mennink, Elmar Tischhauser, and Kan Yasuda. Parallelizable and Authenticated Online Ciphers. In Kazue Sako and Palash Sarkar, editors, *Advances in Cryptology - ASIACRYPT 2013 - Proceedings, Part I*, volume 8269 of *LNCS*, pages 424–443. Springer, 2013.

[Bab89] László Babai. The Fourier Transform and Equations over Finite Abelian Groups : An introduction to the method of trigonometric sums. Lecture notes, December 1989. Available at `http://people.cs.uchicago.edu/~laci/reu02/fourier.pdf`.

[BDJR97] Mihir Bellare, Anand Desai, E. Jokipii, and Phillip Rogaway. A Concrete Security Treatment of Symmetric Encryption. In *Symposium on Foundations of Computer Science - FOCS '97*, pages 394–403. IEEE Computer Society, 1997.

[BDK05] Eli Biham, Orr Dunkelman, and Nathan Keller. Related-Key Boomerang and Rectangle Attacks. In Ronald Cramer, editor, *Advances in Cryptology - EUROCRYPT 2005*, volume 3494 of *LNCS*, pages 507–525. Springer, 2005.

[Ber05] Daniel J. Bernstein. Stronger Security Bounds for Wegman-Carter-Shoup Authenticators. In Ronald Cramer, editor, *Advances in Cryptology - EUROCRYPT 2005*, volume 3494 of *LNCS*, pages 164–180. Springer, 2005.

[BI99] M. Bellare and R. 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*, 1999.

[BK03] Mihir Bellare and Tadayoshi Kohno. A Theoretical Treatment of Related-Key Attacks : RKA-PRPs, RKA-PRFs, and Applications. In Eli Biham, editor, *Advances in Cryptology - EUROCRYPT 2003*, volume 2656 of *LNCS*, pages 491–506. Springer, 2003.

[BKL+07] Andrey Bogdanov, Lars R. Knudsen, Gregor Leander, Christof Paar, Axel Poschmann, Matthew J.B. Robshaw, Yannick Seurin, and Charlotte Vikkelsoe. PRESENT : An Ultra-Lightweight Block Cipher. In Pascal Paillier and Ingrid Verbauwhede, editors, *Cryptographic Hardware and Embedded Systems - CHES 2007*, volume 4727 of *LNCS*, pages 450–466. Springer, 2007.

[BKL+12] Andrey Bogdanov, Lars R. Knudsen, Gregor Leander, François-Xavier Standaert, John P. Steinberger, and Elmar Tischhauser. Key-Alternating Ciphers in a Provable Setting : Encryption Using a Small Number of Public Permutations - (Extended Abstract). In David Pointcheval and Thomas Johansson, editors, *Advances in Cryptology - EUROCRYPT 2012*, volume 7237 of *LNCS*, pages 45–62. Springer, 2012.

[BKR98] M. Bellare, T. Krovetz, and P. Rogaway. Luby-Rackoff Backwards : Increasing Security by Making Block Ciphers Non-invertible. In Kaisa Nyberg, editor, *Advances in cryptology - EUROCRYPT 1998*, pages 266–280. Springer-Verlag, 1998.

[BR93] Mihir Bellare and Phillip Rogaway. Random Oracles are Practical : A Paradigm for Designing Efficient Protocols. In *ACM Conference on Computer and Communications Security*, pages 62–73, 1993.

[BR94] Mihir Bellare and Phillip Rogaway. Optimal Asymmetric Encryption. In Alfredo De Santis, editor, *Advances in Cryptology - EUROCRYPT '94*, volume 950 of *LNCS*, pages 92–111. Springer, 1994.

[BR06a] Mihir Bellare and Phillip Rogaway. The Security of Triple Encryption and a Framework for Code-Based Game-Playing Proofs. In Serge Vaudenay, editor, *Advances in Cryptology - EUROCRYPT 2006*, volume 4004 of *LNCS*, pages 409–426. Springer, 2006. Full version available at `http://eprint.iacr.org/2004/331`.

[BR06b] Mihir Bellare and Phillip Rogaway. The Security of Triple Encryption and a Framework for Code-Based Game-Playing Proofs. In Serge Vaudenay, editor, *Advances in Cryptology - EUROCRYPT 2006*, volume 4004 of *LNCS*, pages 409–426. Springer, 2006. Full version available at `http://eprint.iacr.org/2004/331`.

[CLL+14] Shan Chen, Rodolphe Lampe, Jooyoung Lee, Yannick Seurin, and John P. Steinberger. Minimizing the Two-Round Even-Mansour Cipher. In Juan A. Garay and Rosario Gennaro, editors, *Advances in Cryptology - CRYPTO 2014 (Proceedings, Part I)*, volume 8616 of *LNCS*, pages 39–56. Springer, 2014. Full version available at `http://eprint.iacr.org/2014/443`.

[CLP14] Benoit Cogliati, Rodolphe Lampe, and Jacques Patarin. The indistinguishability of the XOR of k permutations. In *Fast Software Encryption - 21st International Workshop, FSE 2014, London, UK, March 3-5, 2014. Revised Selected Papers*, volume 8540 of *Lecture Notes in Computer Science*, pages 285–302. Springer, 2014.

[CLS15] Benoît Cogliati, Rodolphe Lampe, and Yannick Seurin. Tweaking Even-Mansour Ciphers. In Rosario Gennaro and Matthew Robshaw, editors, *Advances in Cryptology - CRYPTO 2015 - Proceedings, Part I*, volume 9215 of *LNCS*, pages 189–208. Springer, 2015. Full version available at `http://eprint.iacr.org/2015/539`.

[CPS14] Benoit Cogliati, Jacques Patarin, and Yannick Seurin. Security Amplification for the Composition of Block Ciphers : Simpler Proofs and

New Results. In Antoine Joux and Amr M. Youssef, editors, *Selected Areas in Cryptography - SAC 2014*, volume 8781 of *LNCS*, pages 129–146. Springer, 2014.

[Cro00] Paul Crowley. Mercy : A Fast Large Block Cipher for Disk Sector Encryption. In Bruce Schneier, editor, *Fast Software Encryption - FSE 2000*, volume 1978 of *LNCS*, pages 49–63. Springer, 2000.

[CS06] Debrup Chakraborty and Palash Sarkar. A General Construction of Tweakable Block Ciphers and Different Modes of Operations. In Helger Lipmaa, Moti Yung, and Dongdai Lin, editors, *Information Security and Cryptology - Inscrypt 2006*, volume 4318 of *LNCS*, pages 88–102. Springer, 2006.

[CS14] Shan Chen and John Steinberger. Tight Security Bounds for Key-Alternating Ciphers. In Phong Q. Nguyen and Elisabeth Oswald, editors, *Advances in Cryptology - EUROCRYPT 2014*, volume 8441 of *LNCS*, pages 327–350. Springer, 2014. Full version available at `http://eprint.iacr.org/2013/222`.

[CS15a] Benoît Cogliati and Yannick Seurin. Beyond-birthday-bound security for tweakable even-mansour ciphers with linear tweak and key mixing. In *Advances in Cryptology - ASIACRYPT 2015 - 21st International Conference on the Theory and Application of Cryptology and Information Security, Auckland, New Zealand, November 29 - December 3, 2015, Proceedings, Part II*, volume 9453 of *Lecture Notes in Computer Science*, pages 134–158. Springer, 2015.

[CS15b] Benoît Cogliati and Yannick Seurin. On the Provable Security of the Iterated Even-Mansour Cipher against Related-Key and Chosen-Key Attacks. In Elisabeth Oswald and Marc Fischlin, editors, *Advances in Cryptology - EUROCRYPT 2015 - Proceedings, Part I*, volume 9056 of *LNCS*, pages 584–613. Springer, 2015. Full version available at `http://eprint.iacr.org/2015/069`.

[CS16a] Benoît Cogliati and Yannick Seurin. EWCDM : An Efficient, Beyond-Birthday Secure, Nonce-Misuse Resistant MAC. In *Advances in Cryptology - CRYPTO 2016 - Proceedings*, LNCS. Springer, 2016. To appear.

[CS16b] Benoit Cogliati and Yannick Seurin. Strengthening the known-key security notion for block ciphers. In *Fast Software Encryption - 23rd International Workshop, FSE 2016, Bochum, Germany, March 20-23, 2016. Revised Selected Papers*, Lecture Notes in Computer Science. Springer, 2016. To appear.

[DR02] Joan Daemen and Vincent Rijmen. *The Design of Rijndael : AES - The Advanced Encryption Standard*. Springer, 2002.

[EM97] Shimon Even and Yishay Mansour. A Construction of a Cipher from a Single Pseudorandom Permutation. *Journal of Cryptology*, 10(3) :151–162, 1997.

[FLS+10] Niels Ferguson, Stefan Lucks, Bruce Schneier, Doug Whiting, Mihir Bellare, Tadayoshi Kohno, Jon Callas, and Jesse Walker. The Skein Hash Function Family. SHA3 Submission to NIST (Round 3), 2010.

[FP15] Pooya Farshim and Gordon Procter. The Related-Key Security of Iterated Even-Mansour Ciphers. In Gregor Leander, editor, *Fast Software Encryption - FSE 2015*, volume 9054 of *LNCS*, pages 342–363. Springer, 2015. Full version available at `http://eprint.iacr.org/2014/953`.

[FS86] Amos Fiat and Adi Shamir. How to Prove Yourself : Practical Solutions to Identification and Signature Problems. In Andrew M. Odlyzko, editor, *Advances in Cryptology - CRYPTO '86*, volume 263 of *LNCS*, pages 186–194. Springer, 1986.

[GGM86] Oded Goldreich, Shafi Goldwasser, and Silvio Micali. How to construct random functions. *J. ACM*, 33(4) :792–807, 1986.

[GHL+07] David Goldenberg, Susan Hohenberger, Moses Liskov, Elizabeth Crump Schwartz, and Hakan Seyalioglu. On Tweaking Luby-Rackoff Blockciphers. In Kaoru Kurosawa, editor, *Advances in Cryptology - ASIACRYPT 2007*, volume 4833 of *LNCS*, pages 342–356. Springer, 2007.

[GJMN16] Robert Granger, Philipp Jovanovic, Bart Mennink, and Samuel Neves. Improved masking for tweakable blockciphers with applications to authenticated encryption. In *Advances in Cryptology - EUROCRYPT 2016 - 35th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Vienna, Austria, May 8-12, 2016, Proceedings, Part I*, pages 263–293, 2016.

[GM09] Peter Gazi and Ueli M. Maurer. Cascade Encryption Revisited. In Mitsuru Matsui, editor, *Advances in Cryptology - ASIACRYPT 2009*, volume 5912 of *LNCS*, pages 37–51. Springer, 2009.

[GPPR11] Jian Guo, Thomas Peyrin, Axel Poschmann, and Matthew J. B. Robshaw. The LED Block Cipher. In Bart Preneel and Tsuyoshi Takagi, editors, *Cryptographic Hardware and Embedded Systems - CHES 2011*, volume 6917 of *LNCS*, pages 326–341. Springer, 2011.

[HR03] Shai Halevi and Phillip Rogaway. A Tweakable Enciphering Mode. In Dan Boneh, editor, *Advances in Cryptology - CRYPTO 2003*, volume 2729 of *LNCS*, pages 482–499. Springer, 2003.

[HR04] Shai Halevi and Phillip Rogaway. A Parallelizable Enciphering Mode. In Tatsuaki Okamoto, editor, *Topics in Cryptology - CT-RSA 2004*, volume 2964 of *LNCS*, pages 292–304. Springer, 2004.

[HR10] Viet Tung Hoang and Phillip Rogaway. On Generalized Feistel Networks. In Tal Rabin, editor, *Advances in Cryptology - CRYPTO 2010*, volume 6223 of *LNCS*, pages 613–630. Springer, 2010.

[HWKS98] Chris Hall, David Wagner, John Kelsey, and Bruce Schneier. Building PRFs from PRPs. In Hugo Krawczyk, editor, *Advances in Cryptology - CRYPTO '98*, volume 1462 of *LNCS*, pages 370–389. Springer, 1998.

[JNP14] Jérémy Jean, Ivica Nikolic, and Thomas Peyrin. Tweaks and Keys for Block Ciphers : The TWEAKEY Framework. In Palash Sarkar and Tetsu Iwata, editors, *Advances in Cryptology - ASIACRYPT 2014 - Proceedings, Part II*, volume 8874 of *LNCS*, pages 274–288. Springer, 2014.

[JÖS12] Dimitar Jetchev, Onur Özen, and Martijn Stam. Understanding Adaptivity : Random Systems Revisited. In Xiaoyun Wang and Kazue Sako, editors, *Advances in Cryptology - ASIACRYPT 2012*, volume 7658 of *LNCS*, pages 313–330. Springer, 2012.

[Kah96] David Kahn. *The Codebreakers : The Comprehensive History of Secret Communication from Ancient Times to the Internet*. Scribner, rev sub edition, December 1996.

[KPS13] Eike Kiltz, Krzysztof Pietrzak, and Mario Szegedy. Digital Signatures with Minimal Overhead from Indifferentiable Random Invertible Functions. In Ran Canetti and Juan A. Garay, editors, *Advances in Cryptology - CRYPTO 2013 (Proceedings, Part I)*, volume 8042 of *LNCS*, pages 571–588. Springer, 2013.

[Kur10] K. Kurosawa. Power of a public random permutation and its application to authenticated encryption. *IEEE Transactions on Information Theory*, 56(10) :5366–5374, Oct 2010.

[Lee13] Jooyoung Lee. Towards Key-Length Extension with Optimal Security : Cascade Encryption and Xor-cascade Encryption. In Thomas Johansson and Phong Q. Nguyen, editors, *Advances in Cryptology - EUROCRYPT 2013*, volume 7881 of *LNCS*, pages 405–425. Springer, 2013.

[Lin02] T. Lindvall. *Lectures on the Coupling Method*. Dover Books on Mathematics Series. Dover Publications, Incorporated, 2002.

[LPS12] Rodolphe Lampe, Jacques Patarin, and Yannick Seurin. An Asymptotically Tight Security Analysis of the Iterated Even-Mansour Cipher. In Xiaoyun Wang and Kazue Sako, editors, *Advances in Cryptology - ASIACRYPT 2012*, volume 7658 of *LNCS*, pages 278–295. Springer, 2012.

[LR86] Michael Luby and Charles Rackoff. Pseudo-random Permutation Generators and Cryptographic Composition. In *Symposium on Theory of Computing - STOC '86*, pages 356–363. ACM, 1986.

[LR88] Michael Luby and Charles Rackoff. How to Construct Pseudorandom Permutations from Pseudorandom Functions. *SIAM Journal on Computing*, 17(2) :373–386, 1988.

[LRW02] Moses Liskov, Ronald L. Rivest, and David Wagner. Tweakable Block Ciphers. In Moti Yung, editor, *Advances in Cryptology - CRYPTO 2002*, volume 2442 of *LNCS*, pages 31–46. Springer, 2002.

[LS14] Rodolphe Lampe and Yannick Seurin. Security Analysis of Key-Alternating Feistel Ciphers. In Carlos Cid and Christian Rechberger, editors, *Fast Software Encryption - FSE 2014*, volume 8540 of *LNCS*, pages 243–264. Springer, 2014.

[LST12] Will Landecker, Thomas Shrimpton, and R. Seth Terashima. Tweakable Blockciphers with Beyond Birthday-Bound Security. In Reihaneh Safavi-Naini and Ran Canetti, editors, *Advances in Cryptology - CRYPTO 2012*, volume 7417 of *LNCS*, pages 14–30. Springer, 2012. Full version available at `http://eprint.iacr.org/2012/450`.

[Luc00] Stefan Lucks. The Sum of PRPs Is a Secure PRF. In Bart Preneel, editor, *Advances in Cryptology - EUROCRYPT 2000*, volume 1807 of *LNCS*, pages 470–484. Springer, 2000.

[Mau02] Ueli M. Maurer. Indistinguishability of Random Systems. In Lars R. Knudsen, editor, *Advances in Cryptology - EUROCRYPT 2002*, volume 2332 of *LNCS*, pages 110–132. Springer, 2002.

[Men15] Bart Mennink. Optimally Secure Tweakable Blockciphers. In Gregor Leander, editor, *Fast Software Encryption - FSE 2015*, volume 9054 of *LNCS*, pages 428–448. Springer, 2015. Full version available at `http://eprint.iacr.org/2015/363`.

[Men16] Bart Mennink. XPX : Generalized tweakable even-mansour with improved security guarantees. In *Advances in Cryptology - CRYPTO 2016 - Proceedings*, LNCS. Springer, 2016. To appear.

[MI08] Atsushi Mitsuda and Tetsu Iwata. Tweakable Pseudorandom Permutation from Generalized Feistel Structure. In Joonsang Baek, Feng Bao, Kefei Chen, and Xuejia Lai, editors, *ProvSec 2008*, volume 5324 of *LNCS*, pages 22–37. Springer, 2008.

[Min06] Kazuhiko Minematsu. Improved Security Analysis of XEX and LRW Modes. In Eli Biham and Amr M. Youssef, editors, *Selected Areas in Cryptography - SAC 2006*, volume 4356 of *LNCS*, pages 96–113. Springer, 2006.

[Min09] Kazuhiko Minematsu. Beyond-Birthday-Bound Security Based on Tweakable Block Cipher. In Orr Dunkelman, editor, *Fast Software Encryption - FSE 2009*, volume 5665 of *LNCS*, pages 308–326. Springer, 2009.

[MP04] Ueli M. Maurer and Krzysztof Pietrzak. Composition of Random Systems : When Two Weak Make One Strong. In Moni Naor, editor, *Theory of Cryptography Conference - TCC 2004*, volume 2951 of *LNCS*, pages 410–427. Springer, 2004.

[MPR07] Ueli M. Maurer, Krzysztof Pietrzak, and Renato Renner. Indistinguishability Amplification. In Alfred Menezes, editor, *Advances in Cryptology - CRYPTO 2007*, volume 4622 of *LNCS*, pages 130–149. Springer, 2007. Full version available at `http://eprint.iacr.org/2006/456`.

[MRS09] Ben Morris, Phillip Rogaway, and Till Stegers. How to Encipher Messages on a Small Domain. In Shai Halevi, editor, *Advances in Cryptology - CRYPTO 2009*, volume 5677 of *LNCS*, pages 286–302. Springer, 2009.

[MT09] Ueli M. Maurer and Stefano Tessaro. Computational Indistinguishability Amplification : Tight Product Theorems for System Composition. In

Shai Halevi, editor, *Advances in Cryptology - CRYPTO 2009*, volume 5677 of *LNCS*, pages 355–373. Springer, 2009.

[MV00] Shiho Moriai and Serge Vaudenay. On the Pseudorandomness of Top-Level Schemes of Block Ciphers. In Tatsuaki Okamoto, editor, *Advances in Cryptology - ASIACRYPT 2000*, volume 1976 of *LNCS*, pages 289–302. Springer, 2000.

[Mye04] Steven Myers. Black-Box Composition Does Not Imply Adaptive Security. In Christian Cachin and Jan Camenisch, editors, *Advances in Cryptology - EUROCRYPT 2004*, volume 3027 of *LNCS*, pages 189–206. Springer, 2004.

[Pat90] Jacques Patarin. Pseudorandom Permutations Based on the DES Scheme. In Gérard D. Cohen and Pascale Charpin, editors, *EUROCODE '90*, volume 514 of *LNCS*, pages 193–204. Springer, 1990.

[Pat91] Jacques Patarin. New Results on Pseudorandom Permutation Generators Based on the DES Scheme. In Joan Feigenbaum, editor, *Advances in Cryptology - CRYPTO '91*, volume 576 of *LNCS*, pages 301–312. Springer, 1991.

[Pat98] Jacques Patarin. About Feistel Schemes with Six (or More) Rounds. In Serge Vaudenay, editor, *Fast Software Encryption - FSE '98*, volume 1372 of *LNCS*, pages 103–121. Springer, 1998.

[Pat03] 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 *LNCS*, pages 513–529. Springer, 2003.

[Pat04] Jacques Patarin. Security of Random Feistel Schemes with 5 or More Rounds. In Matthew K. Franklin, editor, *Advances in Cryptology - CRYPTO 2004*, volume 3152 of *LNCS*, pages 106–122. Springer, 2004.

[Pat08a] Jacques Patarin. A proof of security in o(2n) for the xor of two random permutations. In *Information Theoretic Security, Third International Conference, ICITS 2008, Calgary, Canada, August 10-13, 2008, Proceedings*, pages 232–248, 2008.

[Pat08b] Jacques Patarin. The "Coefficients H" Technique. In Roberto Maria Avanzi, Liam Keliher, and Francesco Sica, editors, *Selected Areas in Cryptography - SAC 2008*, volume 5381 of *LNCS*, pages 328–345. Springer, 2008.

[Pat10] Jacques Patarin. Security of balanced and unbalanced Feistel Schemes with Linear Non Equalities. IACR Cryptology ePrint Archive, Report 2010/293, 2010. Available at `http://eprint.iacr.org/2010/293`.

[Pat13] Jacques Patarin. Security in $o(2^n)$ for the xor of two random permutations \\ - proof with the standard H technique -. *IACR Cryptology ePrint Archive*, 2013 :368, 2013.

[Pie05a] Krzysztof Pietrzak. Composition Does Not Imply Adaptive Security. In Victor Shoup, editor, *Advances in Cryptology - CRYPTO 2005*, volume 3621 of *LNCS*, pages 55–65. Springer, 2005.

[Pie05b] Krzysztof Pietrzak. *Indistinguishability and Composition of Random Systems*. PhD thesis, ETH Zurich, Switzerland, 2005.

[Pie06] Krzysztof Pietrzak. Composition Implies Adaptive Security in Minicrypt. In Serge Vaudenay, editor, *Advances in Cryptology - EUROCRYPT 2006*, volume 4004 of *LNCS*, pages 328–338. Springer, 2006.

[Pro14] Gordon Procter. A Note on the CLRW2 Tweakable Block Cipher Construction. IACR Cryptology ePrint Archive, Report 2014/111, 2014. Available at `http://eprint.iacr.org/2014/111`.

[RBB03] Phillip Rogaway, Mihir Bellare, and John Black. OCB : A block-cipher mode of operation for efficient authenticated encryption. *ACM Trans. Inf. Syst. Secur.*, 6(3) :365–403, 2003.

[Rog04] Phillip Rogaway. Efficient Instantiations of Tweakable Blockciphers and Refinements to Modes OCB and PMAC. In Pil Joong Lee, editor, *Advances in Cryptology - ASIACRYPT 2004*, volume 3329 of *LNCS*, pages 16–31. Springer, 2004.

[RZ11] Phillip Rogaway and Haibin Zhang. Online Ciphers from Tweakable Blockciphers. In Aggelos Kiayias, editor, *Topics in Cryptology - CT-RSA 2011*, volume 6558 of *LNCS*, pages 237–249. Springer, 2011.

[Sch98] Richard Schroeppel. The Hasty Pudding Cipher. AES submission to NIST, 1998.

[Sha49] Claude Shannon. Communication Theory of Secrecy Systems. *Bell System Technical Journal*, 28(4) :656–715, 1949.

[Sho96] Victor Shoup. On Fast and Provably Secure Message Authentication Based on Universal Hashing. In Neal Koblitz, editor, *Advances in Cryptology - CRYPTO '96*, volume 1109 of *LNCS*, pages 313–328. Springer, 1996.

[STA+14] Yu Sasaki, Yosuke Todo, Kazumaro Aoki, Yusuke Naito, Takeshi Sugawara, Yumiko Murakami, Mitsuru Matsui, and Shoichi Hirose. Minalpher v1. Submission to the CAESAR competition, 2014.

[Ste12] John Steinberger. Improved Security Bounds for Key-Alternating Ciphers via Hellinger Distance. IACR Cryptology ePrint Archive, Report 2012/481, 2012. Available at `http://eprint.iacr.org/2012/481`.

[Ste13] John Steinberger. Counting solutions to additive equations in random sets. arXiv Report 1309.5582, 2013. Available at `http://arxiv.org/abs/1309.5582`.

[Vau98] Serge Vaudenay. Provable Security for Block Ciphers by Decorrelation. In Michel Morvan, Christoph Meinel, and Daniel Krob, editors, *Symposium on Theoretical Aspects of Computer Science, STACS 98*, volume 1373 of *LNCS*, pages 249–275. Springer, 1998.

[Vau99] Serge Vaudenay. Adaptive-Attack Norm for Decorrelation and Super-Pseudorandomness. In Howard M. Heys and Carlisle M. Adams, editors, *Selected Areas in Cryptography - SAC '99*, volume 1758 of *LNCS*, pages 49–61. Springer, 1999.

[Vau03] Serge Vaudenay. Decorrelation : A Theory for Block Cipher Security. *Journal of Cryptology*, 16(4) :249–286, 2003.

[WC81] Mark N. Wegman and Larry Carter. New Hash Functions and Their Use in Authentication and Set Equality. *J. Comput. Syst. Sci.*, 22(3) :265–279, 1981.

[ZMI89] Yuliang Zheng, Tsutomu Matsumoto, and Hideki Imai. On the Construction of Block Ciphers Provably Secure and Not Relying on Any Unproved Hypotheses. In Gilles Brassard, editor, *Advances in Cryptology - CRYPTO '89*, volume 435 of *LNCS*, pages 461–480. Springer, 1989.

Annexe A

Autres résultats d'amplification de sécurité

A.1 Preuve du théorème d'amplification pour la composition d'algorithmes de chiffrement résistant aux attaques ncpa

Dans cette section, nous allons illustrer l'utilité de la formule que nous avons introduite dans le lemme 4 en donnant une preuve élémentaire du théorème d'ε-amplification suivant.

Théorème 11 ([Vau98])**.** *Soient E et F deux algorithmes de chiffrement par blocs ayant le même espace de messages $\mathcal{M}$. Pour tout entier naturel q, on a*

$$\mathbf{Adv}^{\mathrm{ncpa}}_{F\circ E}(q) \leq 2\mathbf{Adv}^{\mathrm{ncpa}}_{E}(q)\mathbf{Adv}^{\mathrm{ncpa}}_{F}(q).$$

Démonstration. Fixons un q-uplet quelconque $\mathbf{x} \in (\mathcal{M})_q$. Par définition de la distance statistique et d'après le lemme 3, on a

$$\begin{aligned}\|\mathsf{p}_{F\circ E,\mathbf{x}} - \mathsf{p}^*\| &= \frac{1}{2}\sum_{\mathbf{y}\in(\mathcal{M})_q} |\mathsf{p}_{F\circ E}(\mathbf{x},\mathbf{y}) - \mathsf{p}^*| \\ &= \frac{1}{2}\sum_{\mathbf{y}\in(\mathcal{M})_q} \left|\sum_{\mathbf{z}\in(\mathcal{M})_q} (\mathsf{p}_E(\mathbf{x},\mathbf{z}) - \mathsf{p}^*)(\mathsf{p}_F(\mathbf{z},\mathbf{y}) - \mathsf{p}^*)\right|.\end{aligned}$$

En utilisant l'inégalité triangulaire, on obtient

$$\begin{aligned}\|\mathsf{p}_{F\circ E,\mathbf{x}} - \mathsf{p}^*\| &\leq \frac{1}{2}\sum_{\mathbf{y}\in(\mathcal{M})_q}\sum_{\mathbf{z}\in(\mathcal{M})_q} |\mathsf{p}_E(\mathbf{x},\mathbf{z}) - \mathsf{p}^*||\mathsf{p}_F(\mathbf{z},\mathbf{y}) - \mathsf{p}^*| \\ &\leq \sum_{\mathbf{z}\in(\mathcal{M})_q} \left(|\mathsf{p}_E(\mathbf{x},\mathbf{z}) - \mathsf{p}^*| \times \frac{1}{2}\sum_{\mathbf{y}\in(\mathcal{M})_q} |\mathsf{p}_F(\mathbf{z},\mathbf{y}) - \mathsf{p}^*|\right) \\ &\leq \sum_{\mathbf{z}\in(\mathcal{M})_q} |\mathsf{p}_E(\mathbf{x},\mathbf{z}) - \mathsf{p}^*| \times \|\mathsf{p}_{F,\mathbf{z}} - \mathsf{p}^*\|.\end{aligned}$$

En utilisant le lemme 2, on en déduit que

$$\begin{aligned}\|\mathsf{p}_{F\circ E,\mathbf{x}} - \mathsf{p}^*\| &\leq \mathbf{Adv}_F^{\mathrm{ncpa}}(q) \sum_{\mathbf{z}\in(\mathcal{M})_q} |\mathsf{p}_E(\mathbf{x},\mathbf{z}) - \mathsf{p}^*| \\ &\leq 2\mathbf{Adv}_F^{\mathrm{ncpa}}(q)\|\mathsf{p}_{E,\mathbf{x}} - \mathsf{p}^*\| \\ &\leq 2\mathbf{Adv}_E^{\mathrm{ncpa}}(q)\mathbf{Adv}_F^{\mathrm{ncpa}}(q).\end{aligned}$$

On en déduit le résultat grâce au lemme 2. □

A.2 Un théorème d'amplification pour la sécurité kpa

Nous allons ici donner une seconde illustration de l'utilité du lemme 4 en prouvant un nouveau résultat d'amplification de sécurité face à une nouvelle classe d'adversaire plus faible que celles que nous avons considéré dans ce manuscrit : les attaques à clair connu, également notées attaques kpa (pour *known plaintext attack*).

On dit qu'un distingueur effectue une attaque à clair connu s'il choisit son $q-$uplet de requêtes directes $\mathbf{x} = (x_1, \ldots, x_q)$ uniformément aléatoirement dans $(\mathcal{M})_q$. Tout comme dans le cas des distingueurs ncpa, il existe une formule exacte caractérisant la résistance d'un algorithme de chiffrement par blocs aux attaques kpa que nous allons énoncer dans le lemme suivant.

Lemme 34 (Sécurité kpa [Pat08b])**.** *Soit E un algorithme de chiffrement par blocs d'espace de messages $\mathcal{M}$. Alors, pour tout entier naturel $q \leq |\mathcal{M}|$, on a*

$$\mathbf{Adv}_E^{\mathrm{kpa}}(q) = \mathsf{p}^* \sum_{\mathbf{x}\in(\mathcal{M})_q} \|\mathsf{p}_{E,\mathbf{x}} - \mathsf{p}^*\|.$$

Démonstration. La preuve est très similaire à celle du lemme 2. Commençons par fixer un distingueur kpa quelconque $\mathcal{D}$. Puisque nous ne considérons que les distingueurs déterministes, $\mathcal{D}$ est complètement caractérisé par sa fonction de décision $\phi_{\mathcal{D}} : (\mathcal{M})_q \times (\mathcal{M})_q \to \{0,1\}$, où $\phi_{\mathcal{D}}(\mathbf{x},\mathbf{y})$ est la sortie de $\mathcal{D}$ lorsqu'il reçoit le q-uplet $\mathbf{y} = (y_1, \ldots, y_q)$ en réponse à ses requêtes choisies uniformément aléatoirement $\mathbf{x} = (x_1, \ldots, x_q)$. Par définition de l'avantage,

$$\begin{aligned}\mathbf{Adv}(\mathcal{D}) = \Big| &\sum_{\mathbf{x}',\mathbf{y}\in(\mathcal{M})_q:\phi_{\mathcal{D}}(\mathbf{x}',\mathbf{y})=1} \Pr\left[K \leftarrow_{\$} \mathcal{K} : E_K(\mathbf{x}) = \mathbf{y} \text{ et } \mathbf{x} = \mathbf{x}'\right] \\ &- \sum_{\mathbf{x}',\mathbf{y}\in(\mathcal{M})_q:\phi_{\mathcal{D}}(\mathbf{x}'\mathbf{y})=1} \Pr\left[P \leftarrow_{\$} \mathsf{Perm}(\mathcal{M}) : P(\mathbf{x}) = \mathbf{y} \text{ et } \mathbf{x} = \mathbf{x}'\right]\Big| \\ = \mathsf{p}^*\Big| &\sum_{\mathbf{x}',\mathbf{y}\in(\mathcal{M})_q:\phi_{\mathcal{D}}(\mathbf{x}',\mathbf{y})=1} \Pr\left[K \leftarrow_{\$} \mathcal{K} : E_K(\mathbf{x}') = \mathbf{y}\right] \\ &- \sum_{\mathbf{x}',\mathbf{y}\in(\mathcal{M})_q:\phi_{\mathcal{D}}(\mathbf{x}',\mathbf{y})=1} \Pr\left[P \leftarrow_{\$} \mathsf{Perm}(\mathcal{M}) : P(\mathbf{x}') = \mathbf{y}\right]\Big|.\end{aligned}$$

En effet, $\mathbf{x}$ est choisi uniformément aléatoirement dans $(\mathcal{M})_q$, indépendamment de la clé dans le monde réel et permutation aléatoire dans le monde idéal, avant le début de l'interaction entre $\mathcal{D}$ et l'oracle de permutation. D'où

$$\begin{aligned}\mathbf{Adv}(\mathcal{D}) &= \mathsf{p}^* \left| \sum_{\mathbf{x}' \in (\mathcal{M})_q} \sum_{\mathbf{y} \in (\mathcal{M})_q : \phi_{\mathcal{D}}(\mathbf{x}', \mathbf{y}) = 1} (\mathsf{p}_E(\mathbf{x}', \mathbf{y}) - \mathsf{p}^*) \right| \\ &\leq \mathsf{p}^* \sum_{\mathbf{x}' \in (\mathcal{M})_q} \left| \sum_{\mathbf{y} \in (\mathcal{M})_q : \phi_{\mathcal{D}}(\mathbf{x}', \mathbf{y}) = 1} (\mathsf{p}_E(\mathbf{x}', \mathbf{y}) - \mathsf{p}^*) \right| \\ &\leq \mathsf{p}^* \sum_{\mathbf{x}' \in (\mathcal{M})_q} \| \mathsf{p}_{E,\mathbf{x}'} - \mathsf{p}^* \|\end{aligned} \tag{A.1}$$

par propriété de la distance statistique entre deux lois de probabilité.

Afin de prouver l'égalité de ces deux quantités, considérons le distingueur $\mathcal{D}$ dont les requêtes sont choisies uniformément aléatoirement et qui renvoie 1 si et seulement si la réponse $\mathbf{y}$ de l'oracle de permutation et le q-uplet $\mathbf{x}$ de requêtes aléatoires vérifient l'inégalité $\mathsf{p}_E(\mathbf{x}, \mathbf{y}) \geq \mathsf{p}^*$. Notons Φ sa fonction de décision. Nous allons prouver que l'avantage de ce distingueur est exactement $\mathsf{p}^* \sum_{x' \in (\mathcal{M})_q} \| \mathsf{p}_{E,\mathbf{x}'} - \mathsf{p}^* \|$. Pour ce faire, nous allons utiliser l'égalité (1.1) :

$$\begin{aligned}\mathsf{p}^* \sum_{x' \in (\mathcal{M})_q} \| \mathsf{p}_{E,\mathbf{x}'} - \mathsf{p}^* \| =& \mathsf{p}^* \sum_{x' \in (\mathcal{M})_q} \left(\sum_{\substack{\mathbf{y} \in (\mathcal{M})_q \\ \mathsf{p}_{E,\mathbf{x}'}(\mathbf{y}) \geq \mathsf{p}^*}} (\mathsf{p}_{E,\mathbf{x}'}(\mathbf{y}) - \mathsf{p}^*) \right) \\ =& \sum_{\substack{\mathbf{x}', \mathbf{y} \in (\mathcal{M})_q \\ \mathsf{p}_{E,\mathbf{x}'}(\mathbf{y}) \geq \mathsf{p}^*}} \left(\mathsf{p}^* \mathsf{p}_{E,\mathbf{x}'}(\mathbf{y}) - (\mathsf{p}^*)^2 \right) \\ =& \left| \sum_{\substack{\mathbf{x}', \mathbf{y} \in (\mathcal{M})_q \\ \Phi(\mathbf{x}', \mathbf{y}) = 1}} (\mathsf{p}^* \mathsf{p}_{E,\mathbf{x}'}(\mathbf{y}) - (\mathsf{p}^*)^2) \right| \\ =& \mathbf{Adv}(\mathcal{D}),\end{aligned}$$

où la dernière égalité découle de l'équation (A.1). On en déduit l'égalité recherchée. □

À présent, concentrons-nous sur le problème de l'amplification de sécurité face aux adversaires kpa. Il est facile de voir, grâce à un exemple simple, que la sécurité kpa ne se comporte pas aussi bien que la sécurité ncpa sous la composition. En effet, considérons la famille G de toutes les permutations de $\mathcal{M}$ pour lesquelles 0 est un point fixe (c'est-à-dire que $\forall g \in G, g(0) = 0$) et voyons la comme un algorithme de chiffrement par blocs. Il existe un distingueur kpa très simple pour cette famille de permutation : il suffit de chercher si la paire $(0, 0)$ est présente parmi les q couples clair/chiffrés et de ne renvoyer 1 que dans ce cas. Bien évidemment, cette famille de permutations est stable sous l'opération de composition. Ainsi, cette attaque aura la même probabilité de succès quelle que soit la longueur de la cascade, lorsque l'on

compose G avec lui-même sous des clés indépendantes. Plus précisément, on a que, pour tout $n \geq 1$, l'avantage de ce distingueur kpa contre G^n vaut au moins

$$\frac{q}{|\mathcal{M}|}\left(1 - \frac{1}{|\mathcal{M}|}\right),$$

ce qui prouve que la composition ne permet pas, en général, d'amplifier la sécurité kpa d'un algorithme de chiffrement par blocs. Cependant, nous avons pu prouver le résultat positif suivant.

Théorème 12 ([CPS14])**.** *Soient E et F deux algorithmes de chiffrement par blocs ayant le même espace de messages $\mathcal{M}$. Pour tout entier naturel $q \leq |\mathcal{M}|$, on a*

$$\mathbf{Adv}^{\mathrm{kpa}}_{F\circ E}(q) \leq 2\mathbf{Adv}^{\mathrm{kpa}}_{E}(q)\mathbf{Adv}^{\mathrm{ncpa}}_{F}(q).$$

Démonstration. On a, d'après le lemme 34

$$\mathbf{Adv}^{\mathrm{kpa}}_{F\circ E}(q) = \mathsf{p}^* \sum_{x\in(\mathcal{M})_q} \|\mathsf{p}_{F\circ E,x} - \mathsf{p}^*\|.$$

D'où, par définition de la distance statistique, et en utilisant le lemme 3,

$$\begin{aligned}
\mathbf{Adv}^{\mathrm{kpa}}_{F\circ E}(q) &= \frac{\mathsf{p}^*}{2} \sum_{\mathbf{x}\in(\mathcal{M})_q} \sum_{\mathbf{y}\in(\mathcal{M})_q} |\mathsf{p}_{F\circ E}(\mathbf{x},\mathbf{y}) - \mathsf{p}^*| \\
&= \frac{\mathsf{p}^*}{2} \sum_{\mathbf{x}\in(\mathcal{M})_q} \sum_{\mathbf{y}\in(\mathcal{M})_q} \Big| \sum_{\mathbf{z}\in(\mathcal{M})_q} (\mathsf{p}_E(\mathbf{x},\mathbf{z}) - \mathsf{p}^*)(\mathsf{p}_F(\mathbf{z},\mathbf{y}) - \mathsf{p}^*) \Big| \\
&\leq \frac{\mathsf{p}^*}{2} \sum_{\mathbf{x}\in(\mathcal{M})_q} \sum_{\mathbf{y}\in(\mathcal{M})_q} \sum_{\mathbf{z}\in(\mathcal{M})_q} |\mathsf{p}_E(\mathbf{x},\mathbf{z}) - \mathsf{p}^*||\mathsf{p}_F(\mathbf{z},\mathbf{y}) - \mathsf{p}^*| \\
&\leq \mathsf{p}^* \sum_{\mathbf{x}\in(\mathcal{M})_q} \sum_{\mathbf{z}\in(\mathcal{M})_q} |\mathsf{p}_E(\mathbf{x},\mathbf{z}) - \mathsf{p}^*| \times \frac{1}{2} \sum_{\mathbf{y}\in(\mathcal{M})_q} |\mathsf{p}_F(\mathbf{z},\mathbf{y}) - \mathsf{p}^*| \\
&\leq \mathsf{p}^* \sum_{\mathbf{x}\in(\mathcal{M})_q} \sum_{\mathbf{z}\in(\mathcal{M})_q} |\mathsf{p}_E(\mathbf{x},\mathbf{z}) - \mathsf{p}^*| \times \|\mathsf{p}_{F,\mathbf{z}} - \mathsf{p}^*\|.
\end{aligned}$$

Ainsi, d'après le lemme 2, on a

$$\begin{aligned}
\mathbf{Adv}^{\mathrm{kpa}}_{F\circ E}(q) &\leq \mathbf{Adv}^{\mathrm{ncpa}}_{F}(q) \times \mathsf{p}^* \sum_{\mathbf{x}\in(\mathcal{M})_q} \sum_{\mathbf{z}\in(\mathcal{M})_q} |\mathsf{p}_E(\mathbf{x},\mathbf{z}) - \mathsf{p}^*| \\
&\leq 2\mathbf{Adv}^{\mathrm{ncpa}}_{F}(q) \times \mathsf{p}^* \sum_{\mathbf{x}\in(\mathcal{M})_q} \|\mathsf{p}_{E,\mathbf{x}} - \mathsf{p}^*\| \\
&\leq 2\mathbf{Adv}^{\mathrm{ncpa}}_{F}(q)\mathbf{Adv}^{\mathrm{kpa}}_{E}(q),
\end{aligned}$$

où, pour la dernière inégalité, on a utilisé le lemme 34. □

Revenons à notre exemple précédent. Nous avons vu que l'auto-composition de la famille de permutations G n'en amplifie pas la sécurité face aux attaques kpa. Le résultat que nous venons de démontrer est cohérent avec cette observation. En effet, il existe une simple attaque ncpa qui distingue G d'une permutation aléatoire avec une probabilité très proche de 1 : il suffit de renvoyer 1 si et seulement si 0 est un point fixe de l'algorithme de chiffrement (dans le cas réel, ceci arrive avec probabilité 1, tandis que dans le cas imaginaire, cette situation se produit avec probabilité $1/|\mathcal{M}|$).

A.3 La composition de 3 algorithmes de chiffrement par blocs

Dans cette section, nous allons mener une analyse plus fine de l'amplification de sécurité obtenue en composant trois algorithmes de chiffrement par blocs. Nous allons obtenir un résultat plus précis que celui obtenu par une application directe du théorème 1 pour $n = 3$.

Théorème 13 ([CPS14])**.** *Soient E, F, et G trois algorithmes de chiffrement par blocs ayant le même espace de messages $\mathcal{M}$ et soit q un entier naturel tel que $q \leq |\mathcal{M}|$. Notons $\varepsilon_E = \mathbf{Adv}^{\mathrm{ncpa}}_E(q)$, $\varepsilon_F = \mathbf{Adv}^{\mathrm{ncpa}}_F(q)$, $\varepsilon_{F^{-1}} = \mathbf{Adv}^{\mathrm{ncpa}}_{F^{-1}}(q)$ et $\varepsilon_{G^{-1}} = \mathbf{Adv}^{\mathrm{ncpa}}_{G^{-1}}(q)$. On a*

$$\mathbf{Adv}^{\mathrm{cca}}_{G\circ F\circ E}(q) \leq \varepsilon_E\varepsilon_F + \varepsilon_E\varepsilon_{G^{-1}} + \varepsilon_{F^{-1}}\varepsilon_{G^{-1}} + \min\{\varepsilon_E\varepsilon_F, \varepsilon_E\varepsilon_{G^{-1}}, \varepsilon_{F^{-1}}\varepsilon_{G^{-1}}\}.$$

Démonstration. Comme dans la preuve du théorème 1, nous allons appliquer le lemme 1 avec $\varepsilon_2 = 0$ et un ensemble de mauvaises transcriptions vide.

Fixons deux q-uplets quelconques $\mathbf{x}, \mathbf{y} \in (\mathcal{M})_q$. La transcription τ correspondant à ces deux q-uplets de messages est

$$\tau = ((x_1, y_1), \ldots, (x_q, y_q)).$$

Nous allons minorer le rapport

$$\frac{\Pr[T_{\mathrm{re}} = \tau]}{\Pr[T_{\mathrm{id}} = \tau]} = \frac{\mathsf{p}_{G\circ F\circ E}(\mathbf{x}, \mathbf{y})}{\mathsf{p}^*}.$$

D'après le lemme 4, on a :

$$\mathsf{p}_{G\circ F\circ E}(\mathbf{x}, \mathbf{y}) = \mathsf{p}^* + \sum_{\mathbf{z},\mathbf{t}\in(\mathcal{M})_q} (\mathsf{p}_E(\mathbf{x}, \mathbf{z}) - \mathsf{p}^*)(\mathsf{p}_F(\mathbf{z}, \mathbf{t}) - \mathsf{p}^*)(\mathsf{p}_G(\mathbf{t}, \mathbf{y}) - \mathsf{p}^*).$$

Comme dans la preuve du théorème 1, nous allons nous concentrer sur l'ensemble des paires $(\mathbf{z}, \mathbf{t}) \in ((\mathcal{M})_q)^2$ de q-uplets de messages tels que

$$(\mathsf{p}_E(\mathbf{x}, \mathbf{z}) - \mathsf{p}^*)(\mathsf{p}_F(\mathbf{z}, \mathbf{t}) - \mathsf{p}^*)(\mathsf{p}_G(\mathbf{t}, \mathbf{y}) - \mathsf{p}^*) < 0.$$

Cet ensemble est constitué des quatre sous-ensembles deux à deux disjoints de $((\mathcal{M})_q)^2$ suivants :

$$\begin{aligned}
A_1 &= \{(\mathbf{z}, \mathbf{t}) \in ((\mathcal{M})_q)^2 : (\mathsf{p}_E(\mathbf{x}, \mathbf{z}) > \mathsf{p}^*) \wedge (\mathsf{p}_F(\mathbf{z}, \mathbf{t}) > \mathsf{p}^*) \wedge (\mathsf{p}_G(\mathbf{t}, \mathbf{y}) < \mathsf{p}^*)\} \\
A_2 &= \{(\mathbf{z}, \mathbf{t}) \in ((\mathcal{M})_q)^2 : (\mathsf{p}_E(\mathbf{x}, \mathbf{z}) > \mathsf{p}^*) \wedge (\mathsf{p}_F(\mathbf{z}, \mathbf{t}) < \mathsf{p}^*) \wedge (\mathsf{p}_G(\mathbf{t}, \mathbf{y}) > \mathsf{p}^*)\} \\
A_3 &= \{(\mathbf{z}, \mathbf{t}) \in ((\mathcal{M})_q)^2 : (\mathsf{p}_E(\mathbf{x}, \mathbf{z}) < \mathsf{p}^*) \wedge (\mathsf{p}_F(\mathbf{z}, \mathbf{t}) > \mathsf{p}^*) \wedge (\mathsf{p}_G(\mathbf{t}, \mathbf{y}) > \mathsf{p}^*)\} \\
A_4 &= \{(\mathbf{z}, \mathbf{t}) \in ((\mathcal{M})_q)^2 : (\mathsf{p}_E(\mathbf{x}, \mathbf{z}) < \mathsf{p}^*) \wedge (\mathsf{p}_F(\mathbf{z}, \mathbf{t}) < \mathsf{p}^*) \wedge (\mathsf{p}_G(\mathbf{t}, \mathbf{y}) < \mathsf{p}^*)\}
\end{aligned}$$

Pour $i = 1, \ldots, 4$, définissons

$$S_i = \sum_{(\mathbf{z},\mathbf{t})\in A_i} (\mathsf{p}_E(\mathbf{x}, \mathbf{z}) - \mathsf{p}^*)(\mathsf{p}_F(\mathbf{z}, \mathbf{t}) - \mathsf{p}^*)(\mathsf{p}_G(\mathbf{t}, \mathbf{y}) - \mathsf{p}^*).$$

Alors, on a

$$\mathsf{p}_{G\circ F\circ E}(\mathbf{x},\mathbf{y}) - \mathsf{p}^* \geq S_1 + S_2 + S_3 + S_4.$$

Nous allons à présent minorer les quantités S_i à tour de rôle. Pour S_1, on a

$$\begin{aligned}
S_1 &= \sum_{\mathbf{z},\mathbf{t}\in(\mathcal{M})_q:\begin{cases}\mathsf{p}_E(\mathbf{x},\mathbf{z})>\mathsf{p}^*\\ \mathsf{p}_F(\mathbf{z},\mathbf{t})>\mathsf{p}^*\\ \mathsf{p}_G(\mathbf{t},\mathbf{y})<\mathsf{p}^*\end{cases} \underbrace{(\mathsf{p}_E(\mathbf{x},\mathbf{z})-\mathsf{p}^*)(\mathsf{p}_F(\mathbf{z},\mathbf{t})-\mathsf{p}^*)}_{>0}\underbrace{(\mathsf{p}_G(\mathbf{t},\mathbf{y})-\mathsf{p}^*)}_{\geq -\mathsf{p}^*}\\
&\geq -\mathsf{p}^* \sum_{\mathbf{z},\mathbf{t}\in(\mathcal{M})_q:\begin{cases}\mathsf{p}_E(\mathbf{x},\mathbf{z})>\mathsf{p}^*\\ \mathsf{p}_F(\mathbf{z},\mathbf{t})>\mathsf{p}^*\\ \mathsf{p}_G(\mathbf{t},\mathbf{y})<\mathsf{p}^*\end{cases} (\mathsf{p}_E(\mathbf{x},\mathbf{z})-\mathsf{p}^*)(\mathsf{p}_F(\mathbf{z},\mathbf{t})-\mathsf{p}^*)\\
&\geq -\mathsf{p}^* \sum_{\mathbf{z},\mathbf{t}\in(\mathcal{M})_q:\begin{cases}\mathsf{p}_E(\mathbf{x},\mathbf{z})>\mathsf{p}^*\\ \mathsf{p}_F(\mathbf{z},\mathbf{t})>\mathsf{p}^*\end{cases} (\mathsf{p}_E(\mathbf{x},\mathbf{z})-\mathsf{p}^*)(\mathsf{p}_F(\mathbf{z},\mathbf{t})-\mathsf{p}^*)\\
&\geq -\mathsf{p}^* \sum_{\mathbf{z}\in(\mathcal{M})_q:\mathsf{p}_E(\mathbf{x},\mathbf{z})>\mathsf{p}^*} \left(|\mathsf{p}_E(\mathbf{x},\mathbf{z})-\mathsf{p}^*| \sum_{\mathbf{t}\in(\mathcal{M})_q:\mathsf{p}_F(\mathbf{z},\mathbf{t})>\mathsf{p}^*} |\mathsf{p}_F(\mathbf{z},\mathbf{t})-\mathsf{p}^*|\right).
\end{aligned}$$

Par propriété de la distance statistique, puis en utilisant le lemme 2, on a

$$\begin{aligned}
S_1 &\geq -\mathsf{p}^* \sum_{\mathbf{z}\in(\mathcal{M})_q:\mathsf{p}_E(\mathbf{x},\mathbf{z})>\mathsf{p}^*} |\mathsf{p}_E(\mathbf{x},\mathbf{z})-\mathsf{p}^*| \times \|\mathsf{p}_{F,\mathbf{z}}-\mathsf{p}^*\|\\
&\geq -\mathsf{p}^*\mathbf{Adv}_F^{\mathrm{ncpa}}(q) \sum_{\mathbf{z}\in(\mathcal{M})_q:\mathsf{p}_E(\mathbf{x},\mathbf{z})>\mathsf{p}^*} |\mathsf{p}_E(\mathbf{x},\mathbf{z})-\mathsf{p}^*|\\
&\geq -\mathsf{p}^*\mathbf{Adv}_F^{\mathrm{ncpa}}(q)\|\mathsf{p}_{E,\mathbf{x}}-\mathsf{p}^*\|\\
&\geq -\mathsf{p}^*\mathbf{Adv}_E^{\mathrm{ncpa}}(q)\mathbf{Adv}_F^{\mathrm{ncpa}}(q).
\end{aligned}$$

De même, pour S_2, on a

$$\begin{aligned}
S_2 &\geq -\mathsf{p}^* \sum_{\mathbf{z},\mathbf{t}\in(\mathcal{M})_q:\begin{cases}\mathsf{p}_E(\mathbf{x},\mathbf{z})>\mathsf{p}^*\\ \mathsf{p}_F(\mathbf{z},\mathbf{t})<\mathsf{p}^*\\ \mathsf{p}_G(\mathbf{t},\mathbf{y})>\mathsf{p}^*\end{cases} (\mathsf{p}_E(\mathbf{x},\mathbf{z})-\mathsf{p}^*)(\mathsf{p}_G(\mathbf{t},\mathbf{y})-\mathsf{p}^*)\\
&\geq -\mathsf{p}^* \sum_{\mathbf{z}\in(\mathcal{M})_q:\mathsf{p}_E(\mathbf{x},\mathbf{z})>\mathsf{p}^*} |\mathsf{p}_E(\mathbf{x},\mathbf{z})-\mathsf{p}^*| \sum_{\mathbf{t}\in(\mathcal{M})_q:\mathsf{p}_G(\mathbf{t},\mathbf{y})>\mathsf{p}^*} |\mathsf{p}_G(\mathbf{t},\mathbf{y})-\mathsf{p}^*|\\
&\geq -\mathsf{p}^*\mathbf{Adv}_E^{\mathrm{ncpa}}(q)\mathbf{Adv}_{G^{-1}}^{\mathrm{ncpa}}(q).
\end{aligned}$$

Enfin, pour S_3, on a

$$\begin{aligned}
S_3 &\geq -\mathsf{p}^* \sum_{\mathbf{z},\mathbf{t}\in(\mathcal{M})_q:\begin{cases}\mathsf{p}_E(\mathbf{x},\mathbf{z})<\mathsf{p}^*\\ \mathsf{p}_F(\mathbf{z},\mathbf{t})>\mathsf{p}^*\\ \mathsf{p}_G(\mathbf{t},\mathbf{y})>\mathsf{p}^*\end{cases} (\mathsf{p}_F(\mathbf{z},\mathbf{t})-\mathsf{p}^*)(\mathsf{p}_G(\mathbf{t},\mathbf{y})-\mathsf{p}^*)\\
&\geq -\mathsf{p}^* \sum_{\mathbf{t}\in(\mathcal{M})_q:\mathsf{p}_G(\mathbf{t},\mathbf{y})>\mathsf{p}^*} |\mathsf{p}_G(\mathbf{t},\mathbf{y})-\mathsf{p}^*| \sum_{\mathbf{z}\in(\mathcal{M})_q:\mathsf{p}_F(\mathbf{z},\mathbf{t})>\mathsf{p}^*} |\mathsf{p}_F(\mathbf{z},\mathbf{t})-\mathsf{p}^*|\\
&\geq -\mathsf{p}^* \sum_{\mathbf{t}\in(\mathcal{M})_q:\mathsf{p}_G(\mathbf{t},\mathbf{y})>\mathsf{p}^*} |\mathsf{p}_G(\mathbf{t},\mathbf{y})-\mathsf{p}^*| \times \|\mathsf{p}_{F^{-1},\mathbf{t}}-\mathsf{p}^*\|
\end{aligned}$$

$$\begin{aligned}
&\geq -\mathsf{p}^* \mathbf{Adv}^{\mathrm{ncpa}}_{F^{-1}}(q) \sum_{\mathbf{t}\in(\mathcal{M})_q : \mathsf{p}_G(\mathbf{t},\mathbf{y})>\mathsf{p}^*} |\mathsf{p}_G(\mathbf{t},\mathbf{y}) - \mathsf{p}^*| \\
&\geq -\mathsf{p}^* \mathbf{Adv}^{\mathrm{ncpa}}_{F^{-1}}(q) \mathbf{Adv}^{\mathrm{ncpa}}_{G^{-1}}(q).
\end{aligned}$$

Dans le cas de S_4, chaque stratégie utilisée pour minorer S_1, S_2, ou S_3 peut être utilisée puisque les trois termes du produit

$$(\mathsf{p}_E(\mathbf{x},\mathbf{z}) - \mathsf{p}^*)(\mathsf{p}_F(\mathbf{z},\mathbf{t}) - \mathsf{p}^*)(\mathsf{p}_G(\mathbf{t},\mathbf{y}) - \mathsf{p}^*)$$

sont strictement négatifs. Par exemple, en suivant la seconde stratégie, on a

$$\begin{aligned}
S_4 = &\sum_{\mathbf{z},\mathbf{t}\in(\mathcal{M})_q : \begin{cases} \mathsf{p}_E(\mathbf{x},\mathbf{z})<\mathsf{p}^* \\ \mathsf{p}_F(\mathbf{z},\mathbf{t})<\mathsf{p}^* \\ \mathsf{p}_G(\mathbf{t},\mathbf{y})<\mathsf{p}^* \end{cases}} \underbrace{(\mathsf{p}_E(\mathbf{x},\mathbf{z}) - \mathsf{p}^*)(\mathsf{p}_G(\mathbf{t},\mathbf{y}) - \mathsf{p}^*)}_{>0} \underbrace{(\mathsf{p}_F(\mathbf{z},\mathbf{t}) - \mathsf{p}^*)}_{<0} \\
&\geq -\mathsf{p}^* \sum_{\mathbf{z},\mathbf{t}\in(\mathcal{M})_q : \begin{cases} \mathsf{p}_E(\mathbf{x},\mathbf{z})<\mathsf{p}^* \\ \mathsf{p}_F(\mathbf{z},\mathbf{t})<\mathsf{p}^* \\ \mathsf{p}_G(\mathbf{t},\mathbf{y})<\mathsf{p}^* \end{cases}} |\mathsf{p}_E(\mathbf{x},\mathbf{z}) - \mathsf{p}^*||\mathsf{p}_G(\mathbf{t},\mathbf{y}) - \mathsf{p}^*| \\
&\geq -\mathsf{p}^* \left(\sum_{\mathbf{z}\in(\mathcal{M})_q : \mathsf{p}_E(\mathbf{x},\mathbf{z})<\mathsf{p}^*} |\mathsf{p}_E(\mathbf{x},\mathbf{z}) - \mathsf{p}^*| \right) \left(\sum_{\mathbf{t}\in(\mathcal{M})_q : \mathsf{p}_G(\mathbf{t},\mathbf{y})<\mathsf{p}^*} |\mathsf{p}_G(\mathbf{t},\mathbf{y}) - \mathsf{p}^*| \right) \\
&\geq -\mathsf{p}^* \|\mathsf{p}_{E,\mathbf{x}} - \mathsf{p}^*\| \|\mathsf{p}_{G^{-1},\mathbf{y}} - \mathsf{p}^*\| \\
&\geq -\mathsf{p}^* \mathbf{Adv}^{\mathrm{ncpa}}_{E}(q) \mathbf{Adv}^{\mathrm{ncpa}}_{G^{-1}}(q).
\end{aligned}$$

Comme les deux autres stratégies peuvent aussi s'appliquer, on a

$$S_4 \geq \max\{-\mathsf{p}^* \varepsilon_E \varepsilon_F, -\mathsf{p}^* \varepsilon_E \varepsilon_{G^{-1}}, -\mathsf{p}^* \varepsilon_{F^{-1}} \varepsilon_{G^{-1}}\}.$$

D'où

$$\begin{aligned}
\frac{\Pr[T_{\mathrm{re}} = \tau]}{\Pr[T_{\mathrm{id}} = \tau]} &= \frac{\mathsf{p}_{G\circ F\circ E}(\mathbf{x},\mathbf{y})}{\mathsf{p}^*} \\
&\geq 1 + \frac{S_1 + S_2 + S_3 + S_4}{\mathsf{p}^*} \\
&\geq 1 - \varepsilon_E\varepsilon_F - \varepsilon_E\varepsilon_{G^{-1}} - \varepsilon_{F^{-1}}\varepsilon_{G^{-1}} - \min\{\varepsilon_E\varepsilon_F, \varepsilon_E\varepsilon_{G^{-1}}, \varepsilon_{F^{-1}}\varepsilon_{G^{-1}}\}.
\end{aligned}$$

On en déduit le résultat grâce au lemme 1, pour $\varepsilon_2 = 0$ et $\Theta_{\mathrm{bad}} = \emptyset$. □

Annexe B

Preuves omises

B.1 Une variante du « sum-capture problem »

Dans cette section, nous prouvons un résultat technique dont nous aurons besoin dans la preuve du lemme 9. Cette preuve utilisera des rappels sur l'analyse de Fourier sur $(\{0,1\}^n, \oplus)$ qui sont développés dans la section 4.4.2. Il s'agit d'un résultat de type « sum-capture », c'est-à-dire qu'il majore la quantité

$$\max_{A,B:|A|=|B|=|Y|} |\{(y,a,b) \in Y \times A \times B : y = a \oplus b\}|$$

pour un sous-ensemble aléatoire Y de $\{0,1\}^n$ (ou plus généralement d'un groupe abélien fini). Ce type de résultat concerne généralement un ensemble Y tiré aléatoirement sans remise [Bab89, Ste13]. Le lemme que nous prouvons considère le cas où Y est tiré aléatoirement *avec* remise, c'est-à-dire que Y est une combinaison avec répétition au lieu d'un ensemble.

Lemme 35. *Soit Y^* une combinaison avec répétition composé de q éléments de $\{0,1\}^n$ choisis uniformément aléatoirement et indépendamment. Pour toute paire de sous-ensembles A et B de $\{0,1\}^n$, soient*

$$\mu(Y^*, A, B) = |\{(y,a,b) \in Y^* \times A \times B \ : \ y = a \oplus b\}|$$

et

$$\mu(Y^*) = \max_{A,B:|A|=|B|=q} \mu(Y^*, A, B).$$

Alors, en supposant $q \geq 1$, on a

$$\Pr\left[\mu(Y^*) \geq \frac{q^3}{2^n} + q\sqrt{3nq}\right] \leq \frac{2}{2^n}.$$

Démonstration. Le lecteur peut se référer à la section 4.4.2 pour les résultats principaux d'analyse de Fourier sur $\mathbb{Z}_2^n$ que nous utilisons. Notre preuve est très similaire à celle du lemme 25, le lecteur intéressé pourra y trouver plus d'informations sur la stratégie suivie.

Soient A et B deux sous-ensembles quelconques de taille q de $\{0,1\}^n$. Pour tout sous-ensemble $S \subset \{0,1\}^n$, on note $1_S : \{0,1\}^n \to \{0,1\}$ la fonction caractéristique de S. Remarquons que, dans Y^*, certaines valeurs peuvent être répétées plusieurs fois. On note alors $\delta_{Y^*} : \{0,1\}^n \to \mathbb{N}$ la fonction qui évalue la multiplicité d'une valeur dans Y^*. Alors on a

$$\begin{aligned}
\mu(Y^*, A, B) &= \sum_{y,a \in \{0,1\}^n} \delta_{Y^*}(y) 1_A(a) 1_B(y \oplus a) \\
&= \sum_{y \in \{0,1\}^n} \delta_{Y^*}(y)(1_A * 1_B)(y) \\
&= 2^n \sum_{\alpha \in \{0,1\}^n} \widehat{\delta_{Y^*}}(\alpha)\widehat{(1_A * 1_B)}(\alpha) \\
&= 2^{2n} \sum_{\alpha \in \{0,1\}^n} \widehat{\delta_{Y^*}}(\alpha)\widehat{1_A}(\alpha)\widehat{1_B}(\alpha) \\
&= 2^{2n}\widehat{\delta_{Y^*}}(0)\widehat{1_A}(0)\widehat{1_B}(0) + 2^{2n} \sum_{\alpha \neq 0} \widehat{\delta_{Y^*}}(\alpha)\widehat{1_A}(\alpha)\widehat{1_B}(\alpha).
\end{aligned}$$

Remarquons que, pour tout sous-ensemble S de $\{0,1\}^n$, on a $\widehat{1_S}(0) = \frac{|S|}{2^n}$ et

$$\begin{aligned}
\widehat{\delta_{Y^*}}(0) &= \frac{1}{2^n} \sum_{x \in \{0,1\}^n} \delta_{Y^*}(x)(-1)^{0 \cdot x} \\
&= \frac{1}{2^n} \sum_{x \in \{0,1\}^n} \delta_{Y^*}(x) \\
&= \frac{q}{2^n}.
\end{aligned}$$

Ainsi

$$\begin{aligned}
\mu(Y^*, A, B) &= \frac{q^3}{2^n} + 2^{2n} \sum_{\alpha \neq 0} \widehat{\delta_{Y^*}}(\alpha)\widehat{1_A}(\alpha)\widehat{1_B}(\alpha) \\
&\leq \frac{q^3}{2^n} + 2^{2n} \sum_{\alpha \neq 0} |\widehat{\delta_{Y^*}}(\alpha)| \cdot |\widehat{1_A}(\alpha)| \cdot |\widehat{1_B}(\alpha)| \\
&\leq \frac{q^3}{2^n} + 2^n \Phi(Y^*) \sum_{\alpha \neq 0} |\widehat{1_A}(\alpha)| \cdot |\widehat{1_B}(\alpha)|,
\end{aligned}$$

où

$$\Phi(Y^*) = \max_{\alpha \neq 0} \left\{ 2^n |\widehat{\delta_{Y^*}}(\alpha)| \right\}.$$

D'après l'inégalité de Cauchy-Schwartz, et en utilisant le fait que, pour tout sous-ensemble $S \subseteq \{0,1\}^n$,

$$\sum_{\alpha \in \{0,1\}^n} |\widehat{1_S}(\alpha)|^2 = \frac{|S|}{2^n},$$

on obtient

$$\mu(Y^*, A, B) \leq \frac{q^3}{2^n} + 2^n \Phi(Y^*) \sqrt{\sum_{\alpha \in \{0,1\}^n} |\widehat{1_A}(\alpha)|^2} \cdot \sqrt{\sum_{\alpha \in \{0,1\}^n} |\widehat{1_B}(\alpha)|^2}$$

$$\leq \frac{q^3}{2^n} + \Phi(Y^*)\sqrt{|A| \cdot |B|}$$
$$\leq \frac{q^3}{2^n} + q \cdot \Phi(Y^*).$$

Puisque cette inégalité est vraie quels que soient les sous-ensembles A et B, il suit que

$$\mu(Y^*) \leq \frac{q^3}{2^n} + q \cdot \Phi(Y^*),$$

et ainsi

$$\Pr\left[\mu(Y^*) \geq \frac{q^3}{2^n} + q\sqrt{3nq}\right] \leq \Pr\left[\Phi(Y^*) \geq \sqrt{3nq}\right].$$

Notons $Y^* = \{y_1, \ldots, y_q\}$ en ordonnant les messages de façon arbitraire. Alors on a

$$\begin{aligned}
\Phi(Y^*) &= \max_{\alpha \neq 0} \left\{ 2^n |\widehat{\delta_{Y^*}}(\alpha)| \right\} \\
&= \max_{\alpha \neq 0} \left\{ \left| \sum_{x \in \{0,1\}^n} \delta_{Y^*}(x)(-1)^{\alpha \cdot x} \right| \right\} \\
&= \max_{\alpha \neq 0} \left\{ \left| \sum_{x \in \{0,1\}^n} \sum_{i=1}^{q} 1_{\{y_i\}}(x)(-1)^{\alpha \cdot x} \right| \right\} \\
&= \max_{\alpha \neq 0} \left\{ \left| \sum_{i=1}^{q} (-1)^{\alpha \cdot y_i} \right| \right\}.
\end{aligned}$$

Pour $\alpha \neq 0$, notons $A_i^{(\alpha)} = (-1)^{\alpha \cdot y_i}$ et $A^{(\alpha)} = \sum_{i=1}^{q} A_i^{\alpha}$. Alors $\Phi(Y^*) = \max_{\alpha \neq 0}\{|A^{(\alpha)}|\}$. La variable aléatoire $A^{(\alpha)}$ est la somme de q variables aléatoires indépendantes $A_i^{(\alpha)}$ telles que $\Pr\left[A_i^{(\alpha)} = 1\right] = \Pr\left[A_i^{(\alpha)} = -1\right] = \frac{1}{2}$. La borne de Chernoff adaptée à ce cas particulier donne, pour tout $a > 0$,

$$\Pr\left[|A^{(\alpha)}| \geq a\right] \leq 2e^{\frac{-a^2}{2q}}.$$

En choisissant $a = \sqrt{3nq} > 0$, on a

$$\Pr\left[|A^{(\alpha)}| \geq \sqrt{3nq}\right] \leq 2e^{\frac{-a^2}{2q}} = 2e^{-3n/2} \leq \frac{2}{2^{2n}}$$

puisque $e^{3/4} \geq 2$. Finalement, on a

$$\begin{aligned}
\Pr\left[\mu(Y^*) \geq \frac{q^3}{2^n} + q\sqrt{3nq}\right] &\leq \Pr\left[\max_{\alpha \neq 0}\{|A^{(\alpha)}|\} \geq \sqrt{3nq}\right] \\
&\leq \sum_{\alpha \neq 0} \Pr\left[|A^{(\alpha)}| \geq \sqrt{3nq}\right] \\
&\leq \frac{2}{2^n}.
\end{aligned}$$

□

B.2 Preuve du lemme 9

Considérons tout d'abord la condition (i). Dans le monde idéal, l'oracle renvoie des messages y uniformément aléatoires et indépendants. Ainsi, l'espérance du nombre de requêtes en collision est inférieure à $q^2/2^n$. Si on note C la variable aléatoire qui compte le nombre de requêtes en collision, alors, d'après l'inégalité de Markov, on a

$$\Pr\left[C \geq \sqrt{q}\right] \leq \frac{q^{3/2}}{2^n}.$$

On considère ensuite la condition (ii). Notons que l'on a

$$\alpha(\tau) = \mu(Y^*, X, X) \leq \mu(Y^*)$$

(voir dans la section B.1 pour la définition de μ). Puisque, dans le monde idéal, Y^* est une combinaison avec répétition de valeurs uniformément indépendantes et aléatoires, alors, d'après le lemme 35, on a

$$\Pr\left[\alpha(\tau) \geq q^3/2^n + q\sqrt{3nq}\right] \leq \Pr\left[\mu(Y^*) \geq q^3/2^n + q\sqrt{3nq}\right] \leq \frac{2}{2^n}.$$

Ainsi, d'après l'inégalité de Boole et puisque $q \geq 2$, on a

$$\Pr\left[T_{\mathrm{id}} \in \Theta_{\mathrm{bad}}\right] \leq \frac{2 + q^{3/2}}{2^n} \leq \frac{2q^{3/2}}{2^n}.$$

B.3 Preuve du lemme 10

Tout d'abord, remarquons que, parmi les $s(s-1)$ paires possibles de requêtes qui ne sont pas en collision, au moins $2\alpha(\tau)$ paires ne vérifient pas les conditions (a) et (b). En effet, par définition d'une bonne transcription (plus précisément, la condition (ii)), il ne peut pas y avoir plus de $\alpha(\tau)$ paires $((x,y),(x',y'))$ telles que $y \oplus x' \in X$. De même, il ne peut pas y avoir plus de $\alpha(\tau)$ paires $((x,y),(x',y'))$ telles que $x \oplus x' \in Y$.

Ainsi, on peut minorer $N(t)$ comme suit :

— on peut choisir $((x_1,y_1),(x'_1,y'_1))$ parmi au moins $s(s-1)-2\alpha(\tau)$ possibilités ;

— une fois que $((x_1,y_1),(x'_1,y'_1))$ est fixé, on peut choisir (x_2,y_2) librement parmi les ($s-2$ possibilités restantes ; ensuite, (x'_2,y'_2) doit être différent de (x_1,y_1), (x'_1,y'_1) et (x_2,y_2), et doivent également être tel que $x'_2 \neq y_2 \oplus y_1 \oplus x'_1$ afin de vérifier la condition (c), et tel que $x'_2 \neq x_2 \oplus x_1 \oplus x'_1$ dans le but de vérifier (d), ce qui retire au plus deux possibilités puisque toutes les requêtes effectuées par le distingueur sont deux à deux distinctes ; ainsi, au final, il y a au moins $(s-5)$ possibilités pour (x'_2,y'_2) ; après avoir retiré les (au plus) $2\alpha(\tau)$ paires de requêtes ne vérifiant pas (a) et (b), il reste au moins $(s-2)(s-5)-2\alpha(\tau)$ possibilités pour la paire $((x_2,y_2),(x'_2,y'_2))$;

— supposons que $((x_1,y_1),(x'_1,y'_1)),\ldots,((x_{i-1},y_{i-1}),(x'_{i-1},y'_{i-1}))$ ont été choisis ; on peut choisir (x_i,y_i) librement parmi les $(s-2i+2)$ possibilités restantes ; ensuite, (x'_i,y'_i) doit être différent de (x_1,y_1), (x'_1,y'_1), $\ldots$, (x_{i-1},y_{i-1}),(x'_{i-1},y'_{i-1}),

et (x_i, y_i) ; de plus, ce message doit être tel que $x'_i \neq y_i \oplus y_j \oplus x'_j$ pour tout $j \in \{1, \ldots, i-1\}$ afin de vérifier la condition (c), et tel que $x'_i \neq x_i \oplus x_j \oplus x'_j$ pour tout $j \in \{1, \ldots, i-1\}$ dans le but de vérifier la condition (d) ; au final, il reste au moins $(s - 4i + 3)$ possibilités pour (x'_i, y'_i) ; après avoir retiré les (au plus) $2\alpha(\tau)$ paires de requêtes qui ne vérifient pas les conditions (a) et (b), il reste au moins $(s - 2i + 2)(s - 4i + 3) - 2\alpha(\tau)$ possibilités pour la paire $(x_i, y_i), (x'_i, y'_i))$.

Puisque nous considérons une combinaison de t paires, c'est-à-dire que l'ordre des éléments n'est pas pris en compte, le nombre $N(t)$ des bons ensembles Σ vaut au moins

$$N(t) \geq \frac{1}{t!} \prod_{i=0}^{t-1} \Big((s-2i)(s-4i-1) - 2\alpha(\tau) \Big).$$

Alors

$$\begin{aligned}
N(t) &\geq \frac{(s)_{2t}}{t!} \prod_{i=0}^{t-1} \frac{(s-2i)(s-4i-1) - 2\alpha(\tau)}{(s-2i)(s-2i-1)} \\
&\geq \frac{(s)_{2t}}{t!} \prod_{i=0}^{t-1} \left(1 - \frac{2si - 4i^2 + 2\alpha(\tau)}{(s-2i)(s-2i-1)} \right) \\
&\geq \frac{(s)_{2t}}{t!} \left(1 - \sum_{i=0}^{t-1} \frac{2si - 4i^2 + 2\alpha(\tau)}{(s-2i)(s-2i-1)} \right) \\
&\geq \frac{(s)_{2t}}{t!} \left(1 - \sum_{i=0}^{t-1} \frac{2si + 2\alpha(\tau)}{(s-2M)^2} \right) \\
&\geq \frac{(s)_{2t}}{t!} \left(1 - \frac{sM^2 + 2\alpha(\tau)M}{(s-2M)^2} \right).
\end{aligned}$$

Notons que, puisque τ est une bonne transcription, alors $\alpha(\tau) \leq q^3/2^n + q\sqrt{3nq}$ et $s \geq q - \sqrt{q}$. De plus, $M = q/2^{n/3}$ et $q\sqrt{q} \leq 2^n$. Ainsi,

$$\begin{aligned}
N(t) &\geq \frac{(s)_{2t}}{t!} \left(1 - \frac{q(M^2 + 2M\sqrt{3nq} + 2Mq^2/2^n)}{(q - \sqrt{q} - 2M)^2} \right) \\
&\geq \frac{(s)_{2t}}{t!} \left(1 - \frac{\frac{q^3}{2^{2n/3}} + \frac{2q^2\sqrt{3nq}}{2^{n/3}} + \frac{2q^4}{2^{4n/3}}}{(q - 3\sqrt{q})^2} \right) \\
&\geq \frac{(s)_{2t}}{t!} \left(1 - \frac{4q}{2^{2n/3}} - \frac{8\sqrt{3nq}}{2^{n/3}} - \frac{8q^2}{2^{4n/3}} \right) \\
&\geq \frac{(s)_{2t}}{t!} \left(1 - \frac{8\sqrt{3nq}}{2^{n/3}} - \frac{12q}{2^{2n/3}} \right),
\end{aligned}$$

où nous avons utilisé le fait que $q \geq 36$, ce qui entraîne le fait que $q - 3\sqrt{q} \geq q/2$.

B.4 Preuve du lemme 11

Rappelons que $|X| = q + t$ et $|Y| = r + t$. Nous allons minorer $N'(t)$ comme suit :

— z_1 doit être tel que $z_1 \notin X$ et $z_1 \oplus u_{1,j} \notin Y$, ce qui laisse au moins $2^n - q - t - q_1(r+t)$ possibilités pour z_1 ;
— une fois z_1 choisi, il reste au moins $2^n - q - t - 1 - q_2(r+t+q_1)$ possibilités pour z_2, puisque z_2 doit être différent de z_1 et de $z_1 \oplus u_{1,j} \oplus u_{2,j'}$ pour tout $j \in \{1, \ldots, q_1\}$ et tout $j' \in \{1, \ldots, q_2\}$; il faut également que $z_2 \notin X$ et $z_2 \oplus u_{2,i} \notin Y$ pour tout $i \in \{1, \ldots, q_2\}$;
— une fois z_1 et z_2 choisis, il reste au moins $2^n - q - t - 2 - q_3(r+t+q_1+q_2)$ possibilités pour z_3, puisque z_3 doit être différent de z_1, z_2, $z_1 \oplus u_{1,j} \oplus u_{3,j'}$ pour tout $j \in \{1, \ldots, q_1\}$ et tout $j' \in \{1, \ldots, q_3\}$, ainsi que de $z_2 \oplus u_{2,j} \oplus u_{3,j'}$ pour tout $j \in \{1, \ldots, q_2\}$ et tout $j' \in \{1, \ldots, q_3\}$; il faut également que $z_3 \notin X$ et $z_3 \oplus u_{3,i} \notin Y$ pour tout $i \in \{1, \ldots, q_3\}$;
— etc.

Ainsi, le nombre de bons uplets $\mathbf{z}$ vaut au moins

$$N'(t) \geq \prod_{i=0}^{r'-1} \left(2^n - q - t - i - q_{i+1}\left(r + t + \sum_{j=1}^{i} q_j\right)\right),$$

comme annoncé.

B.5 Preuve du lemme 12

Afin de minorer le rapport ρ, nous allons avoir besoin de nous débarrasser des requêtes en collision. Notons que, puisque τ est une bonne transcription,

$$d \stackrel{\text{def}}{=} r - s = r' - s' \leq \sum_{i=1}^{r} q_i \leq \sqrt{q}. \tag{B.1}$$

pour $0 \leq t \leq M$.

Nous allons à présent réécrire l'équation (1.22) en séparant les requêtes qui sont en collision de celles qui ne le sont pas comme suit. Pour tout entier naturel t tel que $0 \leq t \leq M$, on a

$$\begin{aligned}
N'(t) &\geq \prod_{i=0}^{r'-1} \left(2^n - q - t - i - q_{i+1}\left(r + t + \sum_{j=1}^{i} q_j\right)\right) \\
&\geq \prod_{i=0}^{s'-1} \left(2^n - q - t - i - q_{i+1}\left(r + t + \sum_{j=1}^{i} q_j\right)\right) \\
&\quad \cdot \prod_{i=s'}^{r'-1} \left(2^n - q - t - i - q_{i+1}\left(r + t + \sum_{j=1}^{i} q_j\right)\right) \\
&\geq \prod_{i=0}^{s'-1} (2^n - q - r - 2i - 2t) \cdot \prod_{i=s'}^{r'-1} (2^n - 2q - 2qq_{i+1}),
\end{aligned}$$

où, pour la dernière inégalité, nous avons utilisé le fait que $q_i = 1$ pour $i \in \{1, \ldots, s'\}$, à cause de l'ordre que nous avons imposé aux requêtes. On remarquera de plus que

$$\frac{\prod_{i=s'}^{r'-1} (2^n - 2q - 2qq_{i+1})}{(2^n)_d} \geq \prod_{i=s'}^{r'-1} \frac{2^n - 2q - 2qq_{i+1}}{2^n}$$

$$\geq 1 - \frac{2qd}{2^n} - \frac{2q\sum_{i=s'}^{r'-1} q_{i+1}}{2^n}$$
$$\geq 1 - \frac{4q\sqrt{q}}{2^n}.$$

Ainsi, pour tout entier naturel t tel que $0 \leq t \leq M$, on a

$$N'(t) \geq \left(1 - \frac{4q\sqrt{q}}{2^n}\right)(2^n)_d \prod_{i=0}^{s'-1} (2^n - q - r - 2i - 2t)\,. \tag{B.2}$$

En rassemblant le lemme 10 ainsi que les équations (1.23) et (B.2), on a

$$\begin{aligned}
\rho &\geq \sum_{0\leq t\leq M} \frac{(2^n)^q N(t)N'(t)}{(2^n)_{q+r-t}} \\
&\geq \left(1 - \frac{8\sqrt{3nq}}{2^{n/3}} - \frac{12q}{2^{2n/3}}\right) \sum_{0\leq t\leq M} \frac{(s)_{2t}(2^n)^q N'(t)}{t!(2^n)_{q+r-t}} \\
&\geq \left(1 - \frac{8\sqrt{3nq}}{2^{n/3}} - \frac{12q}{2^{2n/3}}\right)\left(1 - \frac{4q\sqrt{q}}{2^n}\right) B \\
&\geq \left(1 - \frac{8\sqrt{3nq}}{2^{n/3}} - \frac{12q}{2^{2n/3}} - \frac{4q\sqrt{q}}{2^n}\right) B \\
&\geq \left(1 - \frac{8\sqrt{3nq}}{2^{n/3}} - \frac{16q}{2^{2n/3}}\right) B
\end{aligned} \tag{B.3}$$

où

$$B = \sum_{0\leq t\leq M} \frac{(s)_{2t}(2^n)^q \prod_{i=0}^{s'-1} (2^n - q - r - 2i - 2t)}{t!(2^n - d)_{q+s-t}}.$$

Afin de minorer B, nous allons devoir utiliser une astuce introduite dans [CLL+14] : nous allons exploiter le fait que les termes de cette somme sont proches des valeurs de la fonction de masse correspondant à une loi de probabilité hypergéométrique. Il s'agit d'une loi de probabilité discrète qui décrit la probabilité de k succès parmi n tirages sans remise, dans une population finie de N éléments qui contient exactement « bons » éléments et exactement $N - K$ « mauvais » éléments. La probabilité de tirer exactement k éléments dans l'ensemble des K bons éléments (ce qui correspond à k succès) vaut donc

$$\mathbf{Hyp}_{N,K,n}(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} = \frac{(n)_k (K)_k (N-K)_{n-k}}{k!(N)_n}.$$

L'espérance de la loi de probabilité $\mathbf{Hyp}_{N,K,n}$ vaut nK/N. Ainsi

$$B = \sum_{0\le t\le M} \underbrace{\frac{(s)_{2t}}{(s)_t(s)_t}}_{C} \mathbf{Hyp}_{2^n-q,s,s}(t) \times \underbrace{\frac{(2^n)^q(2^n-q)_s \prod_{i=0}^{s'-1}(2^n-q-r-2i-2t)}{(2^n-d)_{q+s-t}(2^n-q-s)_{s-t}}}_{D}. \quad \text{(B.4)}$$

Nous allons maintenant minorer C et D indépendamment de t avant de minorer B grâce à l'inégalité de Markov. Tout d'abord, puisque τ est une bonne transcription, alors $s \ge q - \sqrt{q} \ge q/2$ et on a, pour tout entier naturel t tel que $0 \le t \le M$,

$$C = \frac{(s)_{2t}}{(s)_t(s)_t} \ge \frac{(s-2M)^{2t}}{s^{2t}} \ge 1 - \frac{4tM}{s} \ge 1 - \frac{8M^2}{q} \ge 1 - \frac{8q}{2^{2n/3}}. \quad \text{(B.5)}$$

Ensuite on a

$$\begin{aligned}
D &= \frac{(2^n)^q(2^n-q)_s \prod_{i=0}^{s'-1}(2^n-q-r-2i-2t)}{(2^n-d)_{q+s-t}(2^n-q-s)_{s-t}} \\
&= \frac{\prod_{i=0}^{s'-1}(2^n-q-r-2i-2t)}{(2^n-q-s-t)_{s-2t}} \cdot \frac{(2^n)^q(2^n-q)_s}{(2^n-d)_{q+s-t}(2^n-q-s)_t} \\
&\ge \prod_{i=0}^{s'-1} \frac{2^n-q-r-2t-2i}{2^n-q-s-t-i} \cdot \frac{(2^n)^q(2^n-q)_{s-t}}{(2^n-d)_{q+s-t}} \\
&\ge \prod_{i=0}^{s'-1} \left(1 - \frac{d+t+i}{2^n-q-s-(t+i)}\right) \cdot \frac{(2^n)^q(2^n-q)_{s-t}}{(2^n-d)_q(2^n-d-q)_{s-t}} \\
&\ge \prod_{i=t}^{s-t-1} \left(1 - \frac{d+i}{2^n-q-s-i}\right) \cdot \frac{(2^n)^q}{(2^n-d)_q} \\
&\ge \prod_{i=t}^{s-t-1} \left(1 - \frac{d+i}{2^n-2s-i}\right) \cdot \prod_{i=0}^{q-1} \left(1 + \frac{d+i}{2^n-d-i}\right) \\
&\ge \prod_{i=0}^{q-1} \left(1 - \frac{d+i}{2^n-2s-i}\right) \cdot \left(1 + \frac{d+i}{2^n-d-i}\right) \\
&\ge \prod_{i=0}^{q-1} \left(1 - \frac{(2s+i)(d+i)}{(2^n-2s-i)(2^n-d-i)}\right) \\
&\ge 1 - \frac{6q^3}{(2^n-3q)(2^n-2q)} \\
&\ge 1 - \frac{24q^3}{2^{2n}}. \quad \text{(B.6)}
\end{aligned}$$

Puisque l'espérance de la loi hypergéométrique $\mathbf{Hyp}_{2^n-q,s,s}$ vaut $\frac{s^2}{2^n-q}$, nous avons

$$\sum_{t>M} \mathbf{Hyp}_{2^n-q,s,s}(t) \leq \frac{s^2}{M(2^n-q)} \leq \frac{2q^2}{M2^n}$$

grâce à l'inégalité de Markov et le fait que $q \leq 2^n/6$. Il suit que

$$\sum_{0\leq t\leq M} \mathbf{Hyp}_{2^n-q,s,s}(t) \geq 1 - \frac{2q^2}{M2^n}.$$

En combinant cette inégalité avec les équations (B.4), (B.5) et (B.6), on obtient

$$\begin{aligned} B \geq& \left(1 - \frac{8q}{2^{2n/3}}\right)\left(1 - \frac{24q^3}{2^{2n}}\right) \sum_{t=0}^{M} \mathbf{Hyp}_{N-q,s,s}(t) \\ \geq& \left(1 - \frac{32q}{2^{2n/3}}\right)\left(1 - \frac{2q^2}{M2^n}\right) \\ \geq& 1 - \frac{32q}{2^{2n/3}} - \frac{2q^2}{M2^n} \geq 1 - \frac{34q}{2^{2n/3}}, \end{aligned} \tag{B.7}$$

où, pour la dernière inégalité, on a utilisé le fait que $M = q/2^{n/3}$. En utilisant les équations (B.3) et (B.7), on a

$$\rho \geq 1 - \frac{8\sqrt{3nq}}{2^{n/3}} - \frac{50q}{2^{2n/3}}, \tag{B.8}$$

comme annoncé.

université PARIS-SACLAY | ÉCOLE DOCTORALE Sciences et technologies de l'information et de la communication (STIC)

Titre : Le schéma d'Even-Mansour paramétrable : preuves de sécurité à l'aide de la technique des coefficients H

Mots clés : cryptographie symétrique, preuves de sécurité, au-delà de la borne des anniversaires, algorithmes de chiffrement par blocs paramétrables, construction d'Even-Mansour, indistinguabilité

Résumé : Les algorithmes de chiffrement par blocs paramétrables constituent une généralisation des algorithmes de chiffrement par blocs classiques qui, en plus d'une clé et d'un message à chiffrer ou déchiffrer, admettent un paramètre additionnel, nommé *tweak* en anglais. Le rôle de ce paramètre additionnel est d'apporter une variabilité à l'algorithme de chiffrement, sans qu'il soit nécessaire de changer la clé ou de garder le *tweak* secret. Ce dernier doit également pouvoir être contrôlé par l'adversaire sans dégradation de la sécurité. Dans cette thèse nous nous intéressons à une classe particulière d'algorithmes de chiffrement par blocs, les algorithmes de chiffrement par blocs à clé alternée. Plus précisément, nous étudions la sécurité du schéma d'Even-Mansour, qui constitue une abstraction de la structure de ces algorithmes dans le modèle de la permutation aléatoire, et cherchons à rendre ce schéma paramétrable tout en conservant de fortes garanties de sécurité. À cette fin, nous introduisons une nouvelle construction générique, baptisée TEM, qui remplace les clés de tours de la construction d'Even-Mansour par une valeur qui dépend de la clé et du *tweak*, et en étudions la sécurité dans deux cas : lorsque le mixage de la clé et du *tweak* est linéaire ou lorsqu'il est très non-linéaire. Nos preuves de sécurité utilisent la technique des coefficients H, introduite par Jacques Patarin dans sa thèse de doctorat, qui permet de transformer des problèmes cryptographiques en problèmes combinatoires sur des groupes finis.

Title : The Tweakable Even-Mansour construction : security proofs with the H-coefficients technique

Keywords : symmetric cryptography, security proofs, beyond the birthday bound, tweakable block ciphers, Even-Mansour construction, indistinguishability

Abstract : Tweakable block ciphers are a generalization of classical block ciphers which, in addition to a key and a plaintext or a ciphertext, take an additionnal parameter called a tweak. The goal of this new parameter is to bring variability to the block cipher without needing to change the key or to keep the tweak secret. The tweak should also be adversarially controllable without sacrificing security. In this thesis we study a particular class of block ciphers, namely key-alternating ciphers. More precisely, we study the security of the Even-Mansour scheme, which is an abstraction of these ciphers in the random permutation model, and seek to bring tweakability to this scheme while keeping strong security guarantees. To this end, we introduce a new generic construction, dubbed TEM, which replaces the round keys from the Even-Mansour construction by a value depending on both the key and the tweak, and study its security in two cases : when the tweak and key mixing is linear or highly non-linear. Our security proofs rely on the H-coefficients technique, a technique introduced by Jacques Patarin in his PhD thesis which transforms cryptographic problems into combinatorial problems in finite groups.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France